Human Inspired Multi-Modal Robot Touch

By

NICHOLAS PESTELL

Department of Engineering Mathematics
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Engineering.

APRIL 2020

Word count: 64,870

Abstract

Humans have exquisite capabilities for dexterous manipulation, demonstrated by our ability to perform complex tasks such as surgical operations, playing the guitar or even peeling an orange. Robots have to progress significantly in order to match human performance in such tasks.

It is clear that humans rely heavily on a sense of touch for dexterous manipulation, as is evident from trying to perform even the most simple of task with cold hands. Therefore, this thesis proposes that: *(i)* touch is an essential sensory modality for robots and *(ii)* human touch is an ideal model on which to base robot tactile perceptual systems.

This thesis demonstrates how aspects of human touch, namely multi-modality, can be applied to tactile sensing in robots. Within a set of artificial sensory channels based on peripheral afferents and their associated encoding for discriminative touch are developed and applied to texture perception. Finally, biomimetic tactile sensors are applied to robot hands to aid grasping.

The success of these techniques provides evidence that human-inspiration is a fruitful approach to endowing robots with a sense of touch and is a concept which should be progressed into the natural extensions of this work.

Dedication and acknowledgements

This thesis is dedicated to my family.

Additional thanks go to Benjamin Ward-Cherrier, Luke Cramphorn, Kirsty Aquilina, John Lloyd, Jasper James and other members of the Tactile Robotics Group at the University of Bristol and the BRL. The support and knowledge provided by this team has been invaluable for the entire process.

I would also like to thank the technical staff at the BRL: Ian Horsfield, Andy Stinchcombe, Gareth Griffiths, Sam Coupland, Julian Potter and Jason Welsby for their help and patience with all practical aspects of my PhD.

Finally, thanks to Nathan Lepora who has been a phenomenal PhD supervisor from start to finish. He has encouraged, supported and guided me to a level far beyond what I had expected.

AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED:.....*Nicholas Pestell* ..... DATE:.....*24/04/2020*.....

Table of Contents

List of Tables

List of Figures

Figure **Page**

CHAPTER 1

INTRODUCTION

Imagine performing a simple task such as tying your shoelaces, threading a needle or removing a screw. Now consider trying to perform such a task under conditions where your sense of touch is limited, for example, with numbness from extreme cold. This comparison demonstrates that tasks for which us humans are normally so adept can become near impossible without the ability to 'feel' what we are doing. Just as human performance in such tasks is vastly improved by a sense of touch, it is reasonable to assume that robots will also improve in-terms of dexterity in similar task, in particular, where humans are to be replaced by robots.

There have been many approaches to endowing robots with a sense of touch over the last two decades [1], yet these are often task specific or limited to single tactile dimensions such as shape [2–4], orientation [5, 6] or hardness [7, 8]. We believe a more holistic approach to artificial touch is required if robots are to achieve human levels of dexterity and tactile control. For example, human touch is comprised of multiple sub-modalities each having evolved to fulfil specific roles [9] and humans employ a range of exploratory procedures for actively stimulating these modalities [10]. Indeed, a new school of thought suggests that these sub-modalities are employed in combination, with complex integration towards the end of the somatosensory pathway [11].

In this thesis we develop techniques in the area of robot touch that are based on a comprehensive overview of the human peripheral somatosensory nervous system. We use an established, optical, biomimetic tactile sensor, the TacTip [12, 13] and advance elements of its design. Specifically, we introduce additional sub-modalities in accordance with the range of tactile channels employed by humans. We also develop novel methods of feature extraction inspired by our understanding of the properties of peripheral tactile afferents in humans and theories of how tactile information is encoded within their response.

The remainder of this introductory chapter is organised as follows: In Section 1.1, we motivate

the work within subsequent chapters. Research questions guiding our investigation are outlined in Section 1.2. Section 1.3 presents a summary of the contributions of this thesis. Section 1.4 provides a list of publications resulting from our research. Finally, the overall thesis structure is presented in Section 1.5.

1.1 Motivation

The natural world has been in existence for 4.5 billion years [14]. Throughout this time, biological systems have evolved exquisite capabilities for interacting with their physical environment, via the vehicle of natural selection [15]. It is no wonder, therefore, that engineers of today are inspired by the natural world when designing artificial systems, a concept known as biomimetics [16]. This concept is exemplified by the work of Leanordo da Vinci in his design of flying machines inspired by the flight of birds [17]. We find the development of robots, too, has benefited from biomimetics [18].

There are many ideas for what constitutes a robot, but one relatively common theme is their practical use for performing tasks which have traditionally been carried out by humans [19]. Many of the areas in which we find it most desirable for robots to be deployed are those involving physical interaction with the environment, such as assisted living [20], healthcare [21] and manufacturing [22]. If the goal is for robots to perform human-like performance in these areas, it is natural to model artificial systems on the relevant aspects of human biology. Humans have evolved a refined sense of touch which is continuously implemented when physically interacting with the environment. This motivates research in the field of artificial touch and demonstrates why biomimicry is such a useful tool.

In particular, we are interested in sub-modalities of human touch that are defined according to a discrete set of mechanoreceptors (the sensory end organs of afferent nerves) [9]. Sub-modalities are leveraged, individually or in combination, for the effective sensation of key aspects of tactile interactions [23], thus enabling human touch to be employed in an expansive range of tactile dimensions. A bio-inspired approach to achieving this in robots is to engineer multi-modal tactile sensors. We define multi-modality, in robot touch, as the ability to perceive discretely different aspects of physical interactions; e.g., vibration vs. a high resolution static pressure field. This does not necessarily have to be with discrete technologies, although this is most common, but it can be achieved with advanced feature engineering or post processing techniques.

Research into robot touch often involves using single sensors [5, 6] or individual tactile fingers [24] to progress technology on the fundamentals of tactile sensing. The overall goal, however, is generally considered to be integration of tactile technologies with robot hands to benefit tasks such as grasping [4] and manipulation [25]. To this end, we also consider how the biomimetic TacTip sensor can be integrated with a robot hand for practical application.

1.2 Research Questions

The broad aims of this thesis are to progress the field of robot touch by presenting a set of methods which are heavily based on theories of human tactile perception. Paying such close attention to human touch has proposed benefits, such as providing a set of tools for comparing robot performance to that of humans, reaching human-like dexterity in robotic systems and serving as a possible tool for expanding our understanding of the natural world. Below are a set of specific research questions which support this overarching aim:

(i) Can we develop an optical tactile sensor that provides tactile modalities, additional to the the TacTip's high-resolution low-speed marker output, for benefit in high-speed scenarios?

(ii) Can we effectively model natural SA-I and FA tactile channels with the TacTip?

(iii) Can we use methods from human studies to better understand the properties of the developed 'artificial afferents' and thus identify suitable tactile cues and encoding methods within these channels?

(iv) Do these methods work in a practical application to robot perception; e.g., texture classification, and how well do the results represent the performance of humans?

(v) Can we adapt the TacTip sensor for application with robot hands and does tactile sensing aid robot grasping?

1.3 Contributions

In working towards answering the research questions above, we have contributed to the field of tactile robotics in the following ways:

- Development of a cheap, dual-modal, optical tactile sensor, based on the TacTip (Chapter 3).
- Demonstration of effective and novel application of established biomimetic perceptual algorithms [26] to extremely low resolution optical tactile images, without the need for image processing. This has broad implications for reducing the cost and size of future technologies (Chapter 3).
- Development of novel feature extraction techniques from the TacTip, modelled on human tactile afferents, and presentation of a method for assessing the likeness of these 'artificial afferents' to their natural counterparts (Chapter 4).
- Identification of viable tactile cues for speed invariant dynamic texture perception based on induced vibrations (Chapters 4 and 5).

- Development of a novel approach to robot texture perception, leveraging conventional machine learning techniques with bio-inspired stimulation and tactile features (Chapter 5).
- Design of tactile fingertips for the integration with an industrial robot hand, The Shadow Modular Grasper [27], and integration of tactile sensing with an existing grasp control framework for the application of grasp adjustment (Chapter 6).

1.4 Publications

The following published works in peer-reviewed journals and conferences have contributed to this thesis, and have been divided into two sections based on my input to each publication (main author and co-author):

Main author

- Pestell, Nicholas, Luke Cramphorn, Fotios Papodopolous and Nathan F. Lepora. "A sense of touch for the Shadow Modular Grasper." IEEE Robotics and Automation Letters 4.2 (2019): 2220-2226.
- Pestell, Nicholas, Nathan F. Lepora. "Texture Perception with a Biomimetic Optical Tactile Sensor." Biomimetic and Biohybrid Systems. Living Machines. Lecture Notes in Computer Science, 10928. Springer, Cham (2018): 365-369.
- Pestell, Nicholas, John Lloyd, Jonathan Rossiter and Nathan F. Lepora. "Dual-modal tactile perception and exploration." IEEE Robotics and Automation Letters 3.2 (2018): 1033-1040.
- Pestell, Nicholas, Benjamin Ward-Cherrier, Luke Cramphorn and Nathan F. Lepora. "Tactile Exploration by Contour Following Using a Biomimetic Fingertip." Biomimetic and Biohybrid Systems. Living Machines. Lecture Notes in Computer Science 9793. Springer, Cham (2016): 485-489.

Co-author

- Elkington, Michael P., et al. "Real time defect detection during composite layup via Tactile Shape Sensing." Science and Engineering of Composite Materials (2019).
- Elkington, Michael P., et al. "Layup end effectors with tactile sensing capabilities." Proceedings of the 4th Symposium on Automated Composite Manufacturing (2019).
- James, Jasper W., Nicholas Pestell and Nathan. F. Lepora, "Slip Detection With a Biomimetic Tactile Sensor." IEEE Robotics and Automation Letters 3.4 (2018): 3340-3346.
- Ward-Cherrier, Benjamin, et al. "The TacTip Family: Soft Optical Tactile Sensors with 3D-Printed Biomimetic Morphologies." Soft Robotics 5.2 (2018): 216-227.

1.5 Thesis Structure

The thesis contains six subsequent chapters, the contents of which are outlined below:

- Chapter 2: Here we review the literature surrounding both human and robot touch. We draw analogies between physiology of human touch and tactile technologies in robots. We also draw analogies between theories of how key tactile dimensions are encoded in afferent response and methods of transduction and feature extraction in tactile robotics. In particular, we focus on the tactile dimension of texture in both human and robot touch as this is a primary application of our methods in subsequent chapters. We also provide a brief review of the current state of the art in tactile sensing for robot hands.

- Chapter 3: We present a novel dual-modal optical tactile sensor with a design based heavily on the TacTip. The sensor has two modes of operation which are inspired by high resolution discriminative touch and high-speed contact detection and reflex action in humans. We demonstrate the capacity for these modalities to be used in combination for autonomous contour following.

- Chapter 4: Novel feature extraction techniques are developed which are modelled on human SA-I and FA primary afferents. We also augment the TacTip with an additional 'vibrational channel' which is proposed as a potential analogue of the natural Pacinian system. We visualise the data from these artificial channels under a range of stimulation conditions with the aim of assessing how similar they are to their natural counterparts as well as trying to gain an understanding for their capacity to be leveraged in a texture discrimination task.

- Chapter 5: This chapter examines the capacity for the transduction and feature extraction methods developed in Chapter 4 to be used for artificial texture discrimination. We employ two types of data (static and dynamic), inspired by theories of human texture perception, and compare different artificial encoding mechanisms (spatial and spatio-temporal) of artificial SA-I and FA afferents, also inspired by human texture perception. Additionally, we consider behaviourally relevant tactile cues of the vibrational channel that can realise speed invariant texture perception and test the performance of these cues as predictors of texture.

- Chapter 6: Here we present the design of a TacTip modification for integration with the Shadow Modular Grasper. We describe how tactile sensing was integrated with a pre-existing control framework and demonstrate the capacity for this system to be leveraged in a grasp adjustment task on a set of three real-world objects.

- Chapter 7: Finally, we summarise the outcomes of each chapter and address the specific research questions stated above. We also discuss limitations to the research and suggest future extensions of the work.

CHAPTER 2

LITERATURE REVIEW

This literature review is separated into two distinct sections: Human Touch in Section 2.1 and Robot Touch in Section 2.2.

Our review of human touch begins with its physiology (Section 2.1.1), where we explain sub-modalities and anatomy, particularly focusing on discriminative touch. We then go on to look at how tactile information is encoded in peripheral afferents (Section 2.1.2), where we explain some key tactile dimensions of discriminative touch and then review the literature surrounding how these dimensions are encoded in primary afferents. We pay particular attention to the dimension of texture as this is a subsequent focus in the thesis (Chapter 5).

In Section 2.2, we consider robot touch in relation to its natural analogue, human touch. For example, we first review sensing technologies (Section 2.2.1) drawing the analogy to natural tactile sub-modalities. We then consider the aforementioned key tactile dimensions of discriminative touch in the context of robotics and highlight similarities of artificial transduction and feature extraction to neural peripheral codes (Section 2.2.2). Again, we focus on the tactile dimension of texture. Finally, we present a short review of the state of the art of tactile sensing for robot hands (Section 2.2.3) since this may be generally considered the goal of robot touch.

2.1 Human Touch

Humans have an exquisite ability to perceive many different aspects of physical stimulation via the somatosensory system. It is a complex modality involving a hierarchy of sub-modalities, varied peripheral codes and cortical computations. It is therefore an active area of research in terms of psychology, neuroscience and physiology.

2.1.1 Physiology of Human Touch

In general, tactile sensation in humans and other animals is facilitated by physical transduction of tactile stimulation through skin and other tissue via sustained pressure change or vibration. Mechanoreceptors are sensory receptors which detect physical stimulation [28] and fire action potentials which are transmitted through the nervous system towards the brain. An action potential is a rapid rising and falling of cell membrane potential which drives a charge flow within the nerve cell itself. This electrical signal is communicated to neighbouring cells via neurotransmitters (chemical signals) which travel across synapses (between nerve cells). If the amount of excitatory neurotransmitter crosses a threshold an action potential is produced in the neighbouring cell [29]. Action potentials are referred to as 'spikes' owing to their abrupt form. Fundamentally, information pertaining to the nature of tactile stimulation is 'encoded' within some temporal aspect or aspects of a time-series of spikes known as a 'spike-train'. These spike-trains may be integrated in a complicated fashion either across populations of neurons (spatial), over time (temporal) or over both dimensions (spatio-temporal) to encode tactile information.

Human touch is comprised of multiple sub-modalities each having evolved for perception of specific aspects of physical interaction [9]. In a broad sense these sub-modalities can be categorised as; *discriminative* - associated with normal classification and inference of external tactile qualities and quantities, mediated by myelinated Aβ afferent nerve fibres [9]; *noxious touch* - painful touch, used to signify dangerous interactions, mediated by myelinated Aδ afferent nerve fibres [31]; *affective* - producing an emotional response, mediated by un-myelinated C afferent nerve fibres, encompasses some forms of pain [32]; and *proprioceptive* - a kinaesthetic awareness of the body [33], mediated by Ia afferent nerve fibres [34]. We are primarily concerned with discriminative touch, as has been the trend within the tactile robotics community to date [35] although we will touch briefly on noxious touch and consider its benefits in robotics, a novel contribution of this thesis.

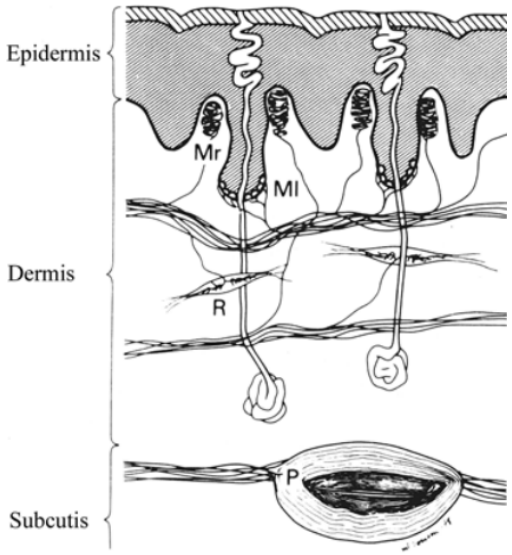

Figure 2.1: Typical schema of mechanoreceptors within glabrous skin of the human hand: Merkel cell complex (Ml), Ruffini endings (R), Meissner's corpuscles (Mr) and Pacinian corpuscles (P). From [30].

2.1.1.1 Discriminative Touch

Discriminative touch is divided further into four channels for perceiving different aspects of physical interactions such as global deformation, stimulus shape and vibrations [23]. These

are defined by afferent type: slowly adapting type I (SA-I), slowly adapting type II (SA-II), fast adapting type I (FA-I or FA) and fast adapting type II (FA-II or PC). All sub-modalities of discriminative touch are mediated by Aβ nerve fibres. This is a broad class of afferent nerve fibres which are myelinated with intermediate to fast conduction velocities (15-100 ms^{-1}) [9].

Adaptation rate refers to how quickly a mechanoreceptor reduces spiking response to sustained stimulation. FA and PC afferents will quickly adapt to a sustained stimulus and thus only fire at the onset or under a changing stimulus. For example, these afferents might continuously respond to vibrations but very quickly stop firing after application of sustained pressure. For this reason, they are sometimes referred to as phasic. In contrast, SA-I and SA-II afferents are tonic [36] meaning they continue to fire for some time under sustained pressure [30]. Fig. 2.2 demonstrates the adaptation rate principle of SA and FA afferents.

The physiological novelty within these sub-modalities arises primarily from their associated mechanoreceptors (the sensory end-organ of Aβ nerve fibres) and their innervation densities within the skin. In some afferent types, mechanoreceptors of a particular class are innervated by single afferent nerve fibres, and in others, afferent fibres branch extensively to innervate multiple mechanoreceptive units. Thus, each afferent corresponds to a particular tactile channel and each afferent has an effective receptive field owing to the arrangement of its associated mechanoreceptors [11]. There are four classes of cutaneous mechanoreceptors found in glabrous (hairless) skin. These are Merkel cells, Ruffini endings, Meissner's corpuscles and Pacinian corpuscles [37], which are innervated by SA-I, SA-II, FA and PC afferents respectively. Fig. 2.1 shows the typical schema for the relative locations and densities of these four mechanoreceptors within glabrous skin of the human hand. We concern ourselves exclusively with sensing in the human hand and specifically fingertips, as these are the primary tools that humans use to explore the world through touch.

SA-I afferents

Numerous Merkel cell units are formed on the end of extensively branched single SA-I afferent axons, and each afferent can innervate as many as 150 Merkel cells [9]. In glabrous skin, Merkel cells are found at the bottom of epidermal ridges (large protrusions of the epidermis into the dermis, see Fig. 2.1) [38]. Studies have demonstrated that the morphology of these ridges could be important for the function of these SA-I afferents: the epidermis is up to 10,000 times stiffer than the underlying dermis, thus the ridges may focus strain energy density propagating through the epidermis [39]. Gerling et al. (2010) analysed two finite element models, with and without intermediate ridges, when undergoing stimulation with gaps and solid indenters and found that forces were concentrated at the tips of the ridges, at the site of Merkel cell complexes, while the no-ridges model diffused strain energy density throughout its underlying tissue [40]. Furthermore, the axis of epidermal ridges align with those of papillary ridges, colloquially known as the fingerprint, and the epidermal ridges follow the movements of papillary ridges [41]; thus the overall structure acts as a mechanical lever for the transmission of touch stimuli to underlying

Merkel cell units. Experiments have estimated that the innervation density of SA-I afferents to the fingertip is ($\sim 80\,\mathrm{cm}^{-2}$) [30] and that they have small, well-defined receptive fields [42]. In combination with psychophysical analysis [43], the size and density of SA-I afferent receptive fields, as well as their adaptation rate, has lead to the hypothesis that SA-I is the receptive afferent for mediating fine surface structures such as textures and shapes and for localising point contacts; i.e., spatial acuity.

SA-II afferents

SA-II afferents innervate Ruffini endings. Unlike SA-I afferents, their axons do not branch, and studies suggest a single SA-II mechanoreceptor supplies each afferent [9]. The spindle shaped receptor resides within the dermis (see Fig. 2.1) [37]. They have a significantly lower density within the fingertip than SA-I afferents ($\sim 10\,\mathrm{cm}^{-2}$) and also have comparatively large receptive fields with soft boundaries [30]. Along with SA-I afferents, they are slowly adapting and in fact can continue to fire up to a minute after the onset of sustained pressure [44]. Unlike SA-I afferents, they are unlikely to be employed for perception of fine spatial detail [44], owing to their low innervation density, large receptive fields and low sensitivity to small deformations [45]. It is believed, instead, that these afferents may be responsible for providing information on skin stretch, shear and relative motion between held objects and the skin [30]. They are seen mostly densely innervating areas of the skin surrounding joints and have been shown to be maximally activated at certain joint angles [46], thus a continuum of these receptors is thought to mediate joint position (proprioception) which is essential for motor control [44].

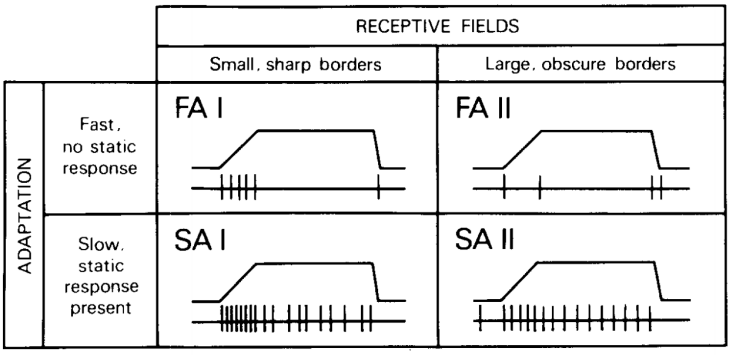

Figure 2.2: The four types of cutaneous mechanoreceptors classified on the basis of their adaptation rates and receptive field types. Figures depict mechanoreceptor activity (bottom) and stimulus pressure profile (top). From [30].

FA afferents

FA afferents innervate Meissner's corpuscles. Each afferent can branch, though not as extensively as SA-I, to innervate multiple Meissner's corpuscles [47]. They are classified as type-I afferents because they share many characteristics with SA-I afferents: they typically have small well defined receptive fields [48] and they provide a high innervation density to the fingertip ($\sim 140\,\mathrm{cm}^{-2}$) [42]; in fact this is higher than SA-I innervation density. These geometric properties have lead many to propose that, along with SA-I afferents, FA could mediate information about

spatial details of skin deformation [30, 42, 44]. These afferents are classed as fast adapting meaning that they are quick to stop firing after the onset of sustained pressure (see Fig. 2.2). They are particularly sensitive to velocity of skin deformation [30] and thus tend to fire whenever skin is deforming. They are particularly sensitive to vibrations in the frequency range between 5 and 40 Hz, often referred to as 'flutter' [49], and are thought to mediate slip information [50]. FA afferents have also been implicated in low threshold contact detection, where single impulses in one or very few afferents is hypothesised to signal contact [51]. Meissner's corpuscles are typically found between epidermal ridges and the smaller limiting ridges [52] (see Fig. 2.1). The purpose of these ridges for FA function is unclear [53], though it has been proposed that the epidermal ridges may amplify deformation aiding acuity, similar to their affect on SA-I function [41].

PC afferents

The sensory end-organs of PC afferents are Pacinian corpuscles (FA-II mechanoreceptors). PC afferents are not branched and terminate with a single Pacinian corpuscle [9]. Studies have estimated the innervation density to the fingertip to be greater than that of SA-II ($\sim 20\,\mathrm{cm}^{-2}$) [42], whilst other experiments have reported PC afferents as the smallest group [44]. In any case, it is known that type-II afferents together constitute a minority innervating the fingertip (30%) [49]. As well as their sparse innervation density PC afferents also share characteristics with SA-II afferents in terms of receptive field size, for which PC afferents have the largest of the four cutaneous afferent types [54]. As with SA-II afferents, they also have receptive fields with obscure borders [30]. Pacinian corpuscles are located deep in subcutis, deeper than the other classes of mechanoreceptors (see Fig. 2.1) which has been offered as an explanation for the large receptive fields of PC afferents [9]. They are the 2^{nd} class of FA afferents and in fact are preferentially sensitive to accelerations and higher order derivatives of skin deformation [30] (see Fig. 2.2). FA and PC afferents can be distinguished in terms of their vibratory thresholds [55]: whilst FA afferents are sensitive to vibrations between 5 and 40 Hz [49], PC afferents show particular sensitivity to much higher-frequency transience, 50 - 500 Hz [44]. At peak frequencies (100-300 Hz) PC afferents have demonstrated extremely low amplitude thresholds (1 μm) [56] and interestingly, at lower frequencies (20 Hz) the response of PC afferents is less than FA and SA counterparts [49]. This has lead to the hypothesis that Pacinian corpuscles mediate information about microtextures during dynamic sliding motions [49], and, indeed, many experiments have found evidence supporting this hypothesis: [57, 58].

It has been proposed that the presence of papillary ridges are critically linked to the role of Pacinian corpuscles in the perception of microtextures [59]. Scheibert et al. [59] found that the presence of an artificial papillary and epidermal ridges on an artificial tactile sensor acted as a band-pass filter, where filtering occurred around a critical frequency which was likely to lie in the range at which Pacinian corpuscles are preferentially sensitive. In comparison, a smooth tactile sensor actually acted as a low pass filter, rapidly attenuating microtexture induced vibrations. The relative amplification of the fingerprint structure was ~100.

Along with FA afferents, PC afferents have also been implicated in signalling low threshold contact detection [51] and also in detecting contact events for lifting and placing objects [44].

2.1.1.2 Proprioceptive Touch

Proprioceptive touch is the system which provides awareness of location of and movement within the body. It is essential for movement as it provides sensory feedback which is used in cortical motor control mechanisms [60]. Proprioception is mediated by proprioceptors which are a class of mechanorecptor located within muscles, tendons and joints [61].

There are three main sensing responsibilities attributed to proprioception: muscle dynamics, limb load and joint position. These a separated according to the mechanisms which enable their sensation. Muscle dynamics (contraction velocity and static length) are detected by receptors which lie within the muscle known as spindles [34]. Contraction velocity is encoded within type Ia afferent fibres and muscle length is encoded within type II afferent fibres [34]. Limb load is sensed via the Golgi-tendon organ, a mechanoreceptor which encodes muscle tension via the Ib afferent fibre [62]. Golgi-tendon organs are comprised of encapsulated fibres located at the interface of muscles and tendons [63]. Joint position is encoded primarily by Ruffini endings [30] which are located within the dermis layer of the skin [37] and are found most densely innervating areas of the skin surrounding joints [46] (for more information on Ruffini endings, see section 2.1.1.1, SA-II afferents).

2.1.1.3 Noxious Touch

Noxious touch is the feeling of pain. Pain is used abundantly in nature to protect animals from harmful situations. Pain pathways have shown exquisite plasticity, a feature which enables humans and animals to learn to avoid danger [31]. There are two major classes of nerve fibres which mediate pain: myelinated Aδ fibres mediate acute, localised 'fast' pain while a sub-set of un-myelinated C fibres mediate 'slow' pain. Both of these fibre types are classed as high threshold mechanoreceptors, meaning that they do not respond to light touch [9]. An acute, potentially localised pain can be experienced when stimulus intensity is raised above a threshold that excites Aδ fibres. When pain is raised further and/or sustained, C fibres will begin to fire, providing the sensation of dull pain [64]. C fibres may also be responsible for pain resulting from inflammation [31]. These two fibre types are classified mainly on the basis of their communication speeds: Aδ fibres transport signals between 5 and 30 ms^{-1} while C fibres transport signal significantly slower at 0.2-2ms-1 [9]. In glabrous skin, both Aδ and C fibres branch at the interface of the dermis and the epidermis and terminate with free nerve endings. It appears common for C fibres to protrude further into the epidermis and terminate shallower in the skin [65].

2.1.1.4 Affective Touch

Affective touch encompasses a subsection of the somatosensory system that elicits an emotional response; e.g., positive emotions, such as love, when experiencing social touch from another human. Affective touch is thought to be an essential tool for social communication [32] and maternal touch has been found to shape the developing "social brain" of an infant [66].

Despite being considered part of the somatosensory system, affective touch is thought to have an entirely distinct neural mechanism. Recent evidence suggests that affective touch is mediated by low threshold mechanorecptive C tactile afferents (CT), in particular, CT afferents are responsive to "slow brush stroking" which is also the socially preferred type of touch [67]. Furthermore, slow brush stroking is shown to activate the insular cortex, an area of the brain which considered as a pathway from sensory systems to the emotional systems of the frontal lobe [68].

2.1.2 Encoding of Tactile Information

In neuroscience it is common to refer to two distinct phases of perception: **encoding** and **decoding** [69]. **Encoding** refers to how stimuli are represented as spike trains in the peripheral and central nervous systems [70]. Conversely **decoding** refers to extracting percepts from neural code; i.e., mapping back to the stimulus. One of the primary goals within computational neuroscience is to develop mathematical models of encoding and decoding which can explain perceptual behaviour in nature [71]. Holdgraf et al. (2017) [72] describes an encoding model as one in which 'stimulus features are used to model brain activity' and a decoding model as one in which 'neural features are used to generated a stimulus output'.

When referring to a single neuron, a stimulus may be encoded with either a **rate code** or a **temporal code**. Central to this distinction is the concept of an *encoding window*, which is the duration of time that the spike train is used to convey information about the relevant stimulus feature or "the duration of a neuron's spike train assumed to correspond to a single symbol within the neural code" (Theunissen and Miller (1995); pg 149 [73]); i.e., the length of spike-train which is required to convey the stimulus. A rate code is one in which relevant stimulus information is "correlated only with the average number of spikes within the encoding window". A temporal code is one in which "information is correlated with some aspect(s) of spikes within the encoding window" [73] (Theunissen and Miller (1995); pg 159 [73]) . This could be precise inter-spike timing or high-frequency firing rate fluctuations [74].

When considering populations of neurons, more abstract encoding mechanisms exist. These are referred to as **intensive** or **spatial**. A stimulus is encoded intensively if the relevant information is contained within firing rate of a single fibre and makes no use of relative locations of responding neurons. Contrastingly, a stimulus is encoded spatially if the relevant information is contained within the spatial modulation of firing rates; i.e., the relative locations of responding neurons is important [43].

Generally, a neural code is referred to as purely spatial if each neuron within the population mediates information in its firing rate only. It is possible that spatial and temporal mechanisms exist in combination (**spatio-temporal code**) [75] where stimulus information is contained in the modulation of spike trains across spatial and temporal dimensions within an encoding window.

It is also useful at this stage to draw the distinction between **cues** and **encoding**. When using the term cue, we are referring to behaviourally relevant properties of the stimulus; i.e., physical properties that are linked to the relevant percept. These may be, for example, *spatial*; e.g., shape of tactile stimulus during static touch, or *temporal*; e.g., frequency of physical vibrations during texture or slip detection. To highlight the distinction from encoding is important because spatial cues may be encoded intensively, spatially or indeed temporally in the spike trains of activated neurons, the same is true of temporal cues. Theunissen and Miller (1995) [73] also draw this distinction, using the term *temporal coding* (as opposed to *temporal encoding*) in reference to where a temporal cue is encoded in neuron firing: "encoding of temporal aspects of a stimulus signal is commonly referred to as temporal coding and could theoretically be implemented through either a rate-encoding scheme or a temporal-encoding scheme".

We also find that, in some cases, **kinaesthetic** mechanisms are hypothesised as an approach of mediating tactile stimuli. These broadly involve using efferent information [76]; i.e., impulses travelling away from the CNS, which command motor control, to encode properties of stimuli involved during haptic exploratory procedures [23]. It also often involves integration of these signals over time in higher order cortical processes in order to perceive traced shape formed by motor commands [77].

2.1.3 Key Tactile Dimensions

Lederman and Klatzky (2009) [23] describe a set of principal tactile dimensions. For touch researchers, it can be helpful to consider individual tactile dimensions in a discrete manner since it is likely that they are encoded, and therefore decoded, in diverse ways. In [23], Lederman and Klatzky make the distinction of 'where' vs. 'what' tactile sensory systems, an idea inspired from vision [78], and separate principal tactile dimensions according to which system they belong. The 'what' system is associated with function and is essentially the process of identification. It encompasses perception of surface properties such as weight, geometry, temperature, compliance, orientation and texture. The 'where' system is associated with guidance of action. It involves perceiving where on the body a stimulus has been applied or perceiving where in space, external to the body, a stimulus is being touched [23]. Lederman and Klatzky provide evidence for the where-what hypothesis, notably in [79], where the authors report that haptic exploration tasks for object recognition activated inferior parietal areas, known for tactile feature integration and naming, whereas haptic exploration tasks for object localisation activated superior parietal regions, known to be used for spatial processing.

Research into encoding of tactile dimensions in humans typically involves psychophysical experiments, where human sensation is measured or recorded in response to controlled tactile stimulation [80] and/or neurophysiological experiments where the activity of peripheral afferents or cortical neurons are measured in mammals such as cats, mice and primates, in response to controlled tactile stimulation. Monkeys, in particular, are believed to have similar neural processing in the somatosensory system as humans, from peripheral to intermediate levels of the central nervous system [81] and their mechanoreceptive afferents are known to have similar innervation densities in the fingertips to that of humans [82]. By comparing how human sensation scales with stimulus dimensions to the scaling of candidate peripheral codes with the identical stimulus dimensions, researchers can gain insight into likely mechanisms of encoding key tactile dimensions.

We will now briefly review psychophysical and neurophysiological literature which aims at understanding encoding of the aforementioned tactile dimensions [10].

We find the tactile dimension of texture to be of particular interest since, as we subsequently outline, it likely depends upon multiple sub-modalities (afferent types) and encoding mechanisms depending upon the scale (coarse-fine) of the stimulus [23]. As such, whilst perhaps being relatively unimportant for manipulation as compared with dimensions of geometry or orientation, we find that it provides a useful representation of the holistic nature of human touch within a single tactile dimension. Therefore, we are interested in studying this aspect of texture perception in robotics and so pay particular attention to literature on human texture perception in the present section.

2.1.3.1 Compliance

Tan et al. (1992, 1993) [83, 84] demonstrated that just noticeable differences (JNDs - defined as the amount that this stimulus intensity must change for the a participant to notice the difference on a defined portion of the trials [85]) for compliance were reduced when participants passively touched stimuli versus active exploring them, alluding to a combination of kinaesthetic and cutaneous modes for compliance perception. Further to this idea, Srinivasan and LaMotte (1995) [86] found that human participants were able to accurately estimate compliance of soft objects using cutaneous touch alone, whereas estimating compliance of a rigid stimulus (spring loaded plate) required the addition of kinaesthetic perception. The authors hypothesise that a spatial code of SA-I and FA firing could be employed for compliance perception when contacting soft objects because the pressure distribution, area of contact and shape of contact interface all depend on object compliance. Conversely, with a rigid stimulus under a given force, the pressure distribution and the geometry of the contact interface will be unaffected by the stiffness of the spring. In a subsequent study LaMotte (2000) [87] examined human performance for discriminating softness through the use of a handheld stylus. As with previous studies, LaMotte found that performance was worse when active touch was not permitted, alluding to the use of

kinaesthetic cues. Interestingly, discrimination was better when participants used tapping as opposed to pressing with the stylus. The authors attribute this to tapping motions enhancing the salience of temporal cues of the inferred change in force; e.g., harder objects produce a greater magnitude and rate of change of force when tapped.

2.1.3.2 Weight

Weber (1834/1978) [88] stated that two modes exist for weight perception, similar to compliance sensing, these are kinaesthetic and cutaneous. A subject of ongoing debate is the degree to which efferent information is employed in the perception of object weight, that is, motor commands sent to muscles during active exploration; e.g., lifting or hefting an object. Waller (1887) [89] found that weight discrimination (defined loosely as the difference in weight the participant was able to perceive with 'certainty') was vastly improved when, opposed to reflex lifting (contraction by 'kathodic excitation'), the participant was allowed to voluntarily lift the stimuli, thus, indicating a bias for efferent information in the discrimination of weight. Similarly, Brodie and Ross (1984) [76] found that discriminative performance (JND) was relatively poor under passive conditions, intermediate under conditions of reflex contraction and best when participants actively lifted stimuli. Brodie et al. were conservative with their interpretation of these results, stating only that "where and how efferent and afferent information are combined to give the weight of the object is not yet known. It would seem improbable, on the basis of the results obtained, that the weight of lifted objects was based solely upon efference copy with no reference to afferent input".

2.1.3.3 Geometry

Geometric properties broadly encompass the shape of an object. More specifically, the curvature of an edge or surface can be considered on two scales: small shapes that can be perceived with a single fingertip and larger shapes which require kinaesthetic inputs realised via haptic explorations such as edge tracing or grasping [23].

For the former, much of the research has been into the specific roles of each type of cutaneous afferent: Goodwin et al. (1997) [90] observed that, in humans, both fast adapting afferents (FA and PC) showed respectively little and no response to indentation with spheres. SA-I and SA-II, however, demonstrated a graded response to curvature, both increasing their firing rates as curvature increased. The authors also found that spatial event plots (SEPs, produced by stimulating each afferent at systematically varied locations across its receptive field [91]), for SA-I afferents only, reflected the curvature of the stimulus. This lead the authors to conclude that SA-I afferents may mediate a spatial code for stimulus curvature by using the liberal interpretation of SEPs [91]: "the single neuron SEP approximates a spatio-temporal neural image of the stimulus as it would appear distributed across a population of similar neurons".

Srinivasan and LaMotte (1987) [92] found that only SA-I afferents responded in a way that can mediate information about gradient of step change in stimulus height. A paired psychophysical

study showed that only SA-I afferent firing could account for the observed discriminability of gradient, meaning that psychophysical performance matched that which would be predicted by a rate code for gradient of step change in SA-I afferents. In a subsequent study [93] the authors found that FA afferents exhibited graded response to bars of varying curvature when dynamically stimulated. The authors summarise the findings of this series of studies thus: "Spatial parameters of primarily SAs in a spatially distributed population of fibres govern the recognition of the overall object shape as a distribution of curvatures; Intensive parameters of only SAs under static indentations, and both SAs and RAs (FAs) under stroking, are important for discriminations of small differences in curvatures of objects belonging to the same category of shape."

For edges with very small radii of curvature, the curvature may relate more significantly to the psychological dimension of sharpness [94] which may be regraded as an important local geometrical feature for shape perception [95]. Kent et al. (2014) [96] suggest that, whilst sharpness may be a specific case of curvature, there may be additional non-spatial mechanisms involved in its perception. This is implied by improved sharpness discrimination when actively exploring edges in the direction parallel with the finger compared with perpendicular exploration. From texture perception studies, parallel exploration is likely to more readily activate FA and PC afferents (thought to encode texture temporally) [59]. This led Kent et al. to suggest that edge sharpness encoding may share similarities with texture encoding when humans employ active exploration.

Kinaesthetic encoding of geometry has been extensively researched over the last 40 years. It is believed that exploration policies guided by aforementioned sensation of local geometry are used, whilst kinaesthetic information of hand and finger movements are integrated in the cortex to form cognition of global shape. Suggestion of this process was made as early as 1978 by Iwamura and Tanaka [77] who observed activation of SI neurons in area 2 of the monkey somatosensory cortex during manipulation. Subsequent experiments involving imaging of human brain activity have provided evidence for this hypothesis [97–100].

2.1.3.4 Temperature

For an in-depth review of temperature encoding and other neurophysiological processes surrounding temperature sensation, we refer the reader to [101]. It is obvious that thermal sensation is important for self preservation: to avoid extremes in either hot or cold. Temperature has also been shown to be useful in discriminative touch: humans are better at discriminating when objects have distinct thermal conductivity [102]. Early research into the encoding of temperature alluded to fibre specificity, where warmth is encoded in un-myelinated C-fibres and cold is encoded in myelinated Aδ-fibres [103–105]. LaMotte and Cambell (1978) [106] and Johnson et al. (1979) [107] compare psychophysical magnitude estimates in humans with peripheral afferent response in 'warm' C-fibres of primates and provide evidence for coding of warmth by simple impulse count

in single fibres and integration of response by a simple additive process in the central nervous system. More recently, Wang et al. (2018) [108] studied hot and cold encoding in root ganglion neurons of anaesthetised mice and found that, whilst warmth was encoded by cumulative impulse as in [107], cold was encoded in a combinatorial fashion, based on co-activation of specific neurons.

2.1.3.5 Orientation

In a tactile context, we refer to the tactile dimension of orientation as the angle of a known object with respect to the skin. For instance, the ability to discriminate the angle of a stimulus edge relative to the fingertip.

An influential study by Johnson and Phillips (1981) [43] was designed to test hypotheses of neural coding of spatial details. In a grating orientation discrimination study, participants were asked to identify relative orientation of two gratings presented in a single trial. Chance-level performance in grating orientation tests was observed at grating periods that match centre-to-centre spacing of SA-I and FA afferents (~1 mm) [42, 109]. In contrast, in two other presented experiments, a two-point discrimination (2PD) test and a gap detection test, above chance-level was observed when two-point spacing and gap width were well bellow 1 mm which authors attribute to varying stimulus areas and edge content (single SA-I and FA afferents have shown intensive sensitivity to edges [112]) in 2PD and gap detection respectively, thus enabling intensive encoding. By extension the suggestion is that a purely spatial code of orientation is leveraged in the grating orientation study.

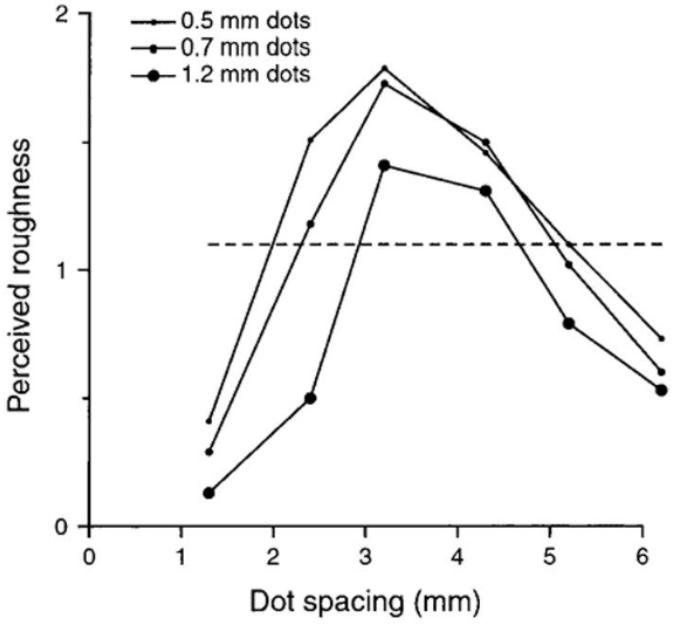

Figure 2.3: Normalised subjectiveive magnitude estimates of roughness for dot spacings ranging from 1.3 to 6.2 mm and dot sizes 0.5, 0.7 and 1.2 mm, made under conditions of passive stimulus scanning under human fingertips. The dashed line shows six examples of stimuli that yielded the same normalised magnitude estimate (from Connor et al. (1990) [110], adapted by Johnson et al. [111]).

Since Johnson and Phillips' study, many experimenters have used grating orientation as a tool for measuring tactile spatial acuity [113, 114]. Notably, Van Boven et al. (2000) [115] used it to show that blind Braille readers exhibit superior tactile spatial acuity over sighted participants .

The idea that stimulus orien-

tation is exclusively mediated by a spatial code is challenged by Pruszynski and Johansson (2014) [116] who demonstrated that single SA-I and FA afferents innervating the human finger can signal edge orientation both intensively and temporally. By recording afferent firing whilst scanning edges of different orientations passed the receptive fields of individual afferents, the authors observed that single afferents showed both intensive and temporal relationships to edge orientation. The authors hypothesised that this was a consequence of non-concentric sensitivity profiles.

2.1.3.6 Texture

Texture is considered a macrostructure, comprised of a consistent arrangement of individual smaller scale shapes. For example, sandpaper is comprised of many tiny glass particles in a roughly periodic arrangement. Properties such as the size and spatial arrangement of these glass particles across the entire piece of sandpaper comprise its texture. On the basis of these properties, human participants may objectively discriminated textures or describe them in terms of subjective dimensions such as roughness or smoothness.

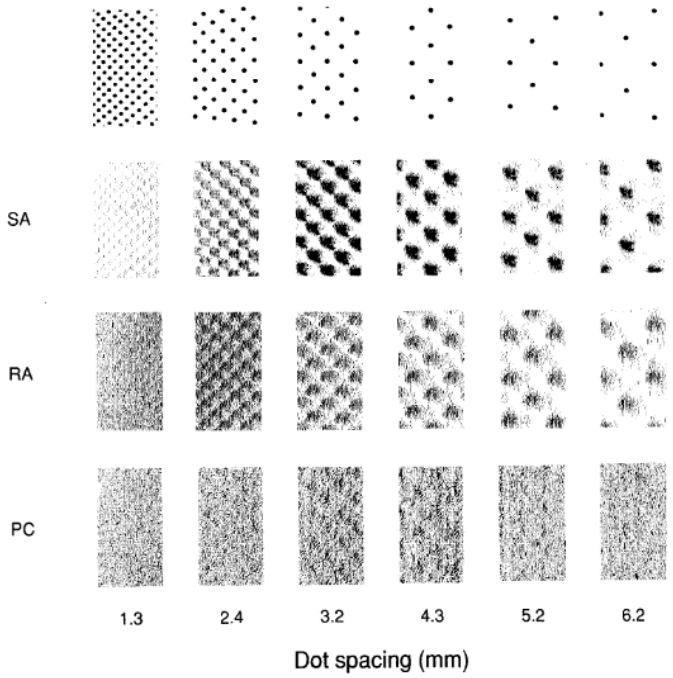

Figure 2.4: Spatial event plots (SEPs) of discharge from a single monkey SA, FA (RA) and PC afferent when passively stimulated with moving dot arrays (top) (from Connor et al. (1990) [110]).

Objective vs. subjective Properties of Texture

It is important to distinguish between objective and subjective textural properties. A participant can distinguish one texture from another, to perceive an objective property of the stimulus. In contrast, such as roughness or smoothness are subjective properties since they do not necessarily relate to objective properties of the stimulus. Johnson et al. (2002) [111] explain that the neural codes which mediate objective and subjective properties are likely to be different. This is succinctly demonstrated in Fig. 2.3, (from Connor et al. (1990) [110], adapted by Johnson et al. [111]) showing that magnitude estimates of roughness increase with dot spacing up to ~3 mm,

above which magnitude estimates begin to fall. Despite the non-monotonic relationship between roughness and dot spacing, human participants are able to distinguish between dot arrays spaced by 2 and 5 mm which have the same perceived roughness (dotted line) [111].

Roughness Estimates of Coarse Textures

Between 1989 and 2001, a number of studies were carried out with the aim of understanding perceptual roughness of relatively coarse textures and in particular the scaling of this subjective dimension with physical stimulus properties such as gap and ridge width of moving stimuli.

For example, Sathian et al. (1989) [117] and Cascio et al. (2001) [118] found that roughness estimates of gratings (spatial periods ~1-4 mm) had a positive relationship to gap width and a negative relationship to ridge width (exemplifying the difference between subjective (roughness) and objective (ridge and gap widths) stimulus properties). By varying the speed of stimulation, Cascio et al. observed that the effect of ridge width could be entirely explained by the associated change in frequency, whereas scanning speed had no effect on the relationship between gap width and roughness, suggesting that the effect of gap width on roughness is explained by purely spatial cues. Both studies found that gap width was more strongly related to roughness than ridge width, indicating that spatial cues dominate subjective roughness estimates on the scales used.

Goodwin et al. (1989) [119] measured the *mean response per grating spatial period* in SA-I, FA and PC afferent fibres innervating a monkey's fingerpad using similar stimuli to those in [117]. Whilst manipulating stimulus speed to maintain consistent frequency, response in all three fibres increased with increasing groove width and remained consistent with changing ridge width. Combined with the psychophysical findings of Cascio et al. (2001) [118], these results indicate that roughness, on the scales used here, may be coded intensively in all three fibre types. Although, importantly, these studies do not provide evidence for or against a spatial code for roughness and indeed the stimuli used do not prohibit an intensive code by virtue of variation in stimulus area.

Unlike the monotonic relationship to spatial period seen by Sathian and Cascio, Connor et al. (1990) [110] found that perceived roughness exhibited an inverted U-shaped function (see Fig. 2.3). This difference may have arisen from the type of texture used: Conner et al. used a tetragonal array of dots rather than the gratings of Sathian and Cascio. Furthermore, the spatial periods used by Connor et al. (~1-6 mm) transcended those used by Sathian or Cascio. Connor et al. also reports results of readings from SA-I, FA and PC peripheral afferents in monkeys under the same stimulation conditions. Unlike Goodwin et al. (1989) [119], they found a simple intensive mean rate code could not explain the relationship between dot spacing and roughness. The authors propose instead a *spatial variation code* which relies on the liberal interpretation of an SEP [91]; i.e., the SEPs shown in Fig. 2.4 are an approximation of a 'moving image' transmitted to the CNS in populations of SA-I and FA afferent fibres. Spatial variation code uses pairs of afferent fibres (positions in the SEP) to provide information about gradients in firing rates across the

skin surface. Spatial variation in SA afferents correlated most closely with perceived roughness followed by FA and finally PC.

Meftah et al. (2000) [120] observed that roughness increased monotonically with spacing between raised dots (1.5-8.5 mm) of consistent diameter and, coherent with the findings of Sathian and Cascio [117, 118], roughness estimates were unaffected by speed (hypothesising that dot spacing is equivalent to gap width and dot diameter to ridge width). Meftah et al. was uncompelled, however, by the *spatial variation code* put forward by Connor et al. [110], suggesting it is unlikely that the "precise and fine-grained" spatial detail is preserved from periphery to SI cortex because this relies on "exquisite point-to-point somatotopy" and evidence suggests that, in fact, the somatosensory pathway converges from periphery to SI cortex [121, 122]. The authors proposed an alternative model to explain speed-invariance of roughness with varying gap width: the signal from a speed only SI neuron is subtracted from the signal of an SI neuron which is sensitive to both speed and spatial period (directly relayed to SI cortex from periphery) thus resulting in a neuron with a graded response to spatial period. Meftah et al. state that SI neurons of these three types have been observed in the SI cortex [123, 124].

Duplex Theory

In his classic text *The World of Touch* (1925) [125], David Katz hypothesised that coarse textures (large distance between microstructures) are readily discerned with spatial characteristics of the stimulus and can be mediated with a static touch, whereas, to discern fine textures, participants must employ a dynamic touch, inducing vibrations, the properties of which provide cues for discriminating said fine texture, an idea which he dubbed 'duplex theory'.

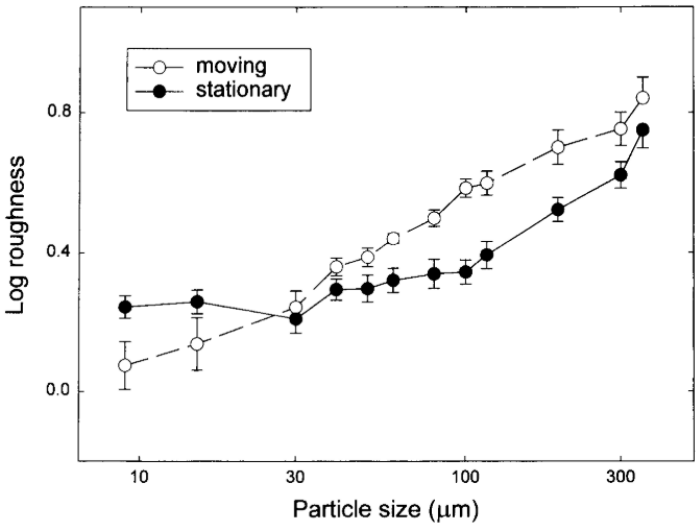

Figure 2.5: Log roughness estimates vs. sandpaper particle size for static touch (filled circles) and dynamic touch (open circles) (from Hollins and Risner (2000) [57]).

Duplex theory is based on the limit presented by the physical resolution of primary afferents in the human finger; i.e., the spatial form of receptive fields. In theory, any observed bound on static texture perception is a consequence of the same process theorised to pose the limit on the aforementioned grating orientation test [43], where chance-level performance was observed at grating separations that matched SA-I and FA afferent spacing (~1 mm) [42, 109].

The evidence of aforementioned experiments into roughness scaling in coarse textures [90, 110, 117, 118, 120] suggests that subjective roughness of moving bumps and gratings, on the scales used, depends, in part, on spatial properties of the stimulus. And the speed-invariance of roughness observed under certain conditions might provide evidence for a *spatial variation code* [110]. However, these studies do not directly allude to the validity of Katz's *duplex theory* [125] primarily because the stimulus scales do not significantly transcend the theoretical spatial resolution of any tactile channel [42, 109].

An influential study by Hollins and Risner (2000) [57] was the first to truly investigate duplex theory. The authors implemented magnitude estimate tests for texture perception of fine (9 μm) to coarse (350 μm) sandpapers under conditions of static touch and passive scanning. The results of this experiment are seen in Fig. 2.5. The gradient of these curves are an indication of the discriminability under both stimulation conditions. It was concluded that for textures with particle sizes <100 μm perception was hindered by the elimination of movement which supports Katz's duplex theory. Interestingly, the coarsest textures used by Hollins and Risner had average spatial periods well below that which could theoretically be perceived with a spatial code predicted by Johnson and Phillips [43]. The authors suggest that, when textural elements range between 100 and 500 μm, rather than information being spatially coded, SA-I firing might be sensitive to total volumetric displacement (intensive encoding) which does, however, still depend on spatial properties of the stimulus. In fact, this agrees with the theory proposed by Johnson and Phillips to explain the disparity in performance between grating orientation and two point discrimination and gap detection experiments [43]. Despite the influence of this work, the authors do, however, concede that their findings fall short of demonstrating that vibrational cues are the primary source of information for fine textures as predicted by Katz.

Perception of Fine Textures

Bensmaia and Hollins (2003) [126] directly investigated the plausibility of vibrations for mediating information about periodic fine surface textures (spatial periods: 16-416 μm). Specifically, they consider two candidate cues which apply to the Pacinain system: *(i)* frequency of vibrations and *(ii)* total spectral power (intensity) weighted by the sensitivity of the Pacinian system. Participants performed free magnitude estimates of roughness during passive scanning whilst the vibrations elicited in the fingertip were simultaneously measured. The effect of changing scanning speed was not significant on perceived roughness as predicted by the frequency based cue, however, roughness estimates were closely matched by the effect of speed on Pacinian-weighted power.

In a follow up study Bensmaia et al. (2005) [127] measured participant's ability to discriminate vibrations delivered by a vibrating platform. The authors found that pairs of stimuli with the same Pacinian-weighted power were readily discriminable suggesting, in fact, that the Pacinian system conveys more than just intensive information about vibration. The authors constructed a model of the Pacinian system which comprised of a set of 'frequency tuned mini-channels'. Spectral dissimilarity between stimulus pairs was computed by performing the sum of the

difference in activation of each mini-channel. The devised spectral model was a better predictor of the psychophysical data, leading the authors to conclude that the Pacinian system simultaneously conveys temporal and intensive information about fine textures. Using a similar experimental arrangement as in [126], Bensmaia and Hollins (2005) [128] found that a similar spectral model as defined in [127] was particularly efficient in predicting the perceived dissimilarity of natural textures, indicating that intensive and temporal features of vibration are important in mediating more complex vibrations.

More recently, Weber et al. (2013) [129] conducted an impressive study which sought to consolidate the findings on texture perception of aforementioned literature. The authors used an extensive set of 55 textures including periodic gratings and dots and some natural surfaces with non-periodic structure. The spatial scales of these textures spanned a range which includes the scales of textures used in all of the aforementioned literature. Readings were taken from SA-I, FA and PC fibres innervating the fingerpads of rhesus macaques whilst stimulating by scanning the textures across their receptive fields. The authors found that SA-I afferents were responsive to coarse but not fine textures and SEPs demonstrated that only textures with spatial periods of greater than ~1 mm elicited a spatial structure indicative of the texture suggesting that SA-I afferents may mediate textural information in relation to the coarsest textures only. FA and PC responses were highly repeatable and temporally patterned and the frequency composition of afferent response matched that of skin vibration. It was found that spike trains of FA and PC afferents convey sufficient information to identify individual textures. The measured spike trains of FA and PC afferents scaled with speed which the authors cite as evidence for roughness perception based on spectral structure of the elicited vibrations because evidence suggests perceived roughness is invariant to speed [120, 130]. A parallel psychophysical study was conducted where participants performed magnitude estimates of roughness under the same stimulation used for afferent recordings. The *spatial variation code* devised by Conner et al. (1990) [110], suggested that spatial variation in SA-I response was a relatively poor predictor of roughness (R^2=0.64 and 0.35 for coarse and fine textures respectively). Temporal variation, however, yielded good results for FA and PC afferents, R^2=0.88 and 0.76 respectively. Importantly, good performance was still observed when the coarsest textures were removed.

speed-invariance of Fine Texture Perception

We have seen that coarse textures are likely encoded spatially in populations of SA-I and FA afferents and that these spatial representations are robust to scanning speed [110]. Likewise, the effect of scanning speed is not significant on perceived roughness of fine textures [126].

Boundy-singer et al. (2017) [131] investigated the degree of speed-invariance of texture perception of natural surfaces. Participants were passively stimulated by 24 natural surfaces, presented at 4 different scanning speeds (40, 80, 120 and 160 mms^{-1}). Participants were asked to rate pairs of presented textures in terms of dissimilarity. In agreement with previous experiments, speed had no significant effect on perceived dissimilarity. The authors propose two competing models

for the observed speed-invariance of texture perception: *(i)* Speed is corrected for by dividing the frequency of each component by a perceived scanning speed. This hypothesis assumes speed can be accurately signalled a in neuron response which is not known; *(ii)* A more compelling theory is that cortical computations are capable of extracting harmonic structure from neuron firing. This idea is inspired by auditory timbre invariance, where it has been shown that, whilst scanning speed shifts frequency composition of spike response, the harmonic structure is preserved [132], thus harmonic structure is a viable code for texture [133]. By harmonic structure, we mean the relative frequencies of fundamentals, harmonics and other significant parts of the signal; i.e. the relative positions of peaks and troughs in the frequency spectra. The overall shape of the frequency spectra (includes amplitudes) is not necessarily invariant since harmonics may move in and out of filter bands depending on the scanning speed.

Texture is clearly a complex tactile dimension: literature suggests that it requires the use of multiple tactile channels (SA-I, FA and PC) and numerous neural codes have been theorised with complex relationships to stimulus scale and nature of contact (static, passive or active). To summarise, the overarching idea is that spatial properties of coarse stimuli are encoded in modulation across populations of SA-I and FA afferents (spatial code) whether this is with a static [43] or dynamic touch [110, 117, 118, 120]. Fine textures have microgeometry which is too small (<~1 mm) to be resolved spatially. Information about fine texture is mediated in vibrations, requiring dynamic touch, [126] and is likely coded in the harmonic structure of FA and PC afferent firing [129]. Some evidence also suggests an intermediary mechanism of intensive encoding in SA-I afferents relying on overall skin displacement resulting from static touch for textures with microstructures in the range of 100 and 500 μm [57]. Texture sensing exhibits perceptual constancy with speed regardless of the stimulus scale and therefore the encoding mechanism [131].

2.2 Robot Touch

Put simply, robot touch is a sensory modality providing robots with the ability to gain an understanding of the palpable world through physical interaction.

Although touch has clear benefits for robots, engineers have a long way to go before this sensory channel provides the level of practicality offered by other modalities; e.g., vision and audition. It seems possible that this discrepancy is a consequence of a lack of tactile sensing hardware, the development of which is in its relative infancy; e.g., in vision, photographic cameras, which can near perfectly capture the visual world as viewed by the human eye, were first developed in 19th century [134], whereas the earliest tactile sensors were in the 1970s [135]. Another plausible explanation is the complexity of its natural analogue. We have seen that within human touch, there exists a number of discriminative sub-modalities of different afferent types (Section 2.1.1.1), each thought to mediate varying stimulus properties; e.g., shape [90], skin

Figure 2.6: Prominent tactile sensors. *Left* to *right*: iCub fingertip, TakkTile, TacTip, GelSight and BioTac.

stretch [30] and vibration [126]. Furthermore, it is believed that a number of peripheral neural codes; e.g., intensive, spatial and temporal (Section 2.1.2), are leveraged in combination with relevant sub-modalities depending on the tactile dimension being perceived. In many cases, the nature of this encoding is an active area of research.

2.2.1 Technologies

In general, a tactile sensor must transduce its physical deformation to a representative electronic signal such that the signal can be decoded to produce an understanding of the stimulus. A vast range of transduction technologies have been explored in the past and, at present, the prominent tactile sensors operate with contrasting principles [1, 136]. This again attests to the idea that, as a sensory modality, touch is an actively developing research field: a single technology, shown to surpass all others, has thus far proven elusive.

When selecting a transduction technology the engineer must carefully consider which tactile dimension they wish to sense: hardware trade-offs commonly include spatial resolution, sensitivity, bandwidth and hysteresis (as well as more practical considerations such as cost and size). It is such properties that lead us to draw the analogy between transduction technologies and natural tactile sub-modalities of discriminative touch. For example, a technology offering high resolution may be ideal for transducing information about detailed spatial structures as is the case with SA-I afferents in human touch [112], owing to its high innervation density and localised receptive fields. Likewise, a technology offering a high sample rate would be considered analogous to the PC channel due to its sensitivity to high frequency vibration [127] and may therefore be ideal for perceiving fine textures via dynamic touch.

Capacitive Sensors

Capacitive sensors consist of two oppositely charged electrodes separated by a compressible dielectric. As the sensor experiences changes in external pressure the gap between the electrodes changes, thus altering the capacitance which is measured [1]. Capacitive technologies are often used to construct small, high bandwidth array, based devices with moderate resolution [137, 138]. One such sensor, which has found wide application is the iCub fingertip [139, 140] (see Fig. 2.6). This sensor is comprised of 12 small capacitive taxels integrated on a single device providing a moderate sample rate of 50 Hz. The sensor is surrounded by a soft 3-layer composite fabric

wrapping and is roughly the shape and size of a human fingertip. The PR2 humanoid robot provides an example of another common capacitive based tactile sensor [141]: a flat 3x5 array of capacitive taxels is located on each finger of the robot's two-fingered grippers. Each tactile fingertip is considerably larger than that of the iCub and thus the spatial resolution is lower. Additionally, the sample rate is approximately half that of the iCub tactile fingertip.

The taxel arrays of the iCub fingertip and PR2 robot gripper are limited to measuring normal forces, whereas tangential shear is an important aspect of many tactile interactions. A capacitive based tactile sensor has been demonstrated to measure tangential shear with a creative physical design of individual silicone "nibs" each associated with one electrode in a 6x6 array [142]. This sensor could additionally measure at 300 Hz, which, in combination with shear sensing capabilities, make it an ideal candidate for slip detection.

Capacitive technologies are often a more expensive option and therefore have often been employed in commercial tactile sensors where the demand is for small form factor. In general, drawbacks associated with capacitive technologies are hysteresis, susceptibility to electronic noise and a non-linear response.

Piezoresistive Sensors

Piezoresistive sensors work by measuring the change in resistance of a deformable semiconducting material. The flexible semiconductor has an associated band-gap which is modified under stress and results in a variable resistance [143]. An early example of the use of resistive technologies in touch was demonstrated by Shimizu et al. (2002) [144]: a primitive sensor was developed for detecting object compliance using a single piezoresistive chip. More sophisticated array based piezoresistive tactile sensors have been developed for compliance sensing; Drimus et al. (2011,2014) [145, 145] and tactile servoing; Li et al. (2013) [5], where the sensor was capable of transducing information about edge orientation, encoded in the spatial arrangement of taxel activation (see Fig. 2.11).

Piezoresistive technologies are one of the best candidates for fabricating tactile 'skin' either by screen printing on flexible substrates [146, 147] or fabrication as small microelectromechanical systems (MEMS) embedded onto flexible circuit boards [148]. These MEMS devices can also provide high sample rate (~300 Hz). For example, in the work of Oddo et al. (2011) [149] where resistive devices were used for detecting vibration in texture perception experiments.

As with capacitive technologies, piezoresistive methods do not naturally enable shear sensing, however, similar to [142] shear sensing was enabled by fabricating physical "cantilevers" of piezoresistive material [150]. Also, piezoresistive sensors tend to exhibit non-linearity, particularly due to changes in temperature [151].

Piezoelectric Sensors

The piezoelectric effect is the accumulation of charge in certain materials when participant to mechanical stress. Stress causes re-orientating of molecular dipole moments which has a net effect on the total polarization. The piezoelectric effect is seen in quartz and man-made ceramics and polymers. The generated charges are directly proportional to the applied mechanical stress [152]. An often-cited benefit of these devices is their excellent high-frequency response [153], which makes them an ideal candidate for detecting vibrotactile stimuli [154]; e.g., for fine textures or slip events [155]. The majority of flexible piezoelectric sensors are made from polymer called polyvinylidene difluoride (PVDF) [155–157] and are commonly used for medical applications such as minimally invasive surgery [158–160], owing to their flexibility and small form factor and have also found application in tactile skin [161]. Piezoelectric materials, however, exhibit response drift as the piezoelectric effect becomes less sensitie over time [162]. They are also sensitive to temperature, since elevated temperatures cause an additional drop in internal resistance and sensitivity, thus temperature diodes are sometimes incorporated for calibration [163]. A fundamental limitation of these technologies is that they are only applicable to dynamic sensing; i.e., they do not respond to static load or deformation [35].

Barometric Sensors

We class barometric sensors as tactile devices which leverage cheap commercially-available MEMS pressure sensing units (barometers) [164]. At its core, the individual MEMS barometric device is generally piezoresistive: a silicon diaphragm deforms under changing pressure and the output resistance is measured. In tactile sensors, sustained pressure or vibration at the contact interface propagate through a medium and is measured by the MEMS device. Typically the pressure sensing device is embedded within a deformable silicone which aids propagation of pressure waves and also provides a soft surface which is amenable to interactions such as grasping and manipulation. A commonly used barometric tactile sensor is the TAkkTile [165] (see Fig. 2.6) which consists of a set MEMS pressure sensors on a PCB strip embedded within PDMS silicone, this particular sensor has seen application with robot hands [166]. Barometric tactile sensors have also been incorporated with flexible circuit boards for the development of tactile skin [167]. The main benefits of these sensors are high sample rate [1] and extremely low cost. Conversely, a significant drawback of is poor spatial resolution owing to the size of each individual unit. Recently, however, ingenuitive post processing has enabled superresolution, allowing acuity to transcend physical resolution [164]. These devices are also generally of relatively low accuracy and suffer from drift and temperature sensitivity (owing to their piezoresistve basis) and critically are unable to measure shear. However, barometric sensors may be an effective solution for low-cost applications.

Optical Sensors

Optical tactile sensors encompass a range of sensor technologies with the common feature that

light is used to transduce tactile information. For example, fibre-optics [168, 169], waveguides [170] or, more typically, photosensitive arrays (cameras) [171, 172]. In general, the camera-based tactile sensor operates by capturing images of a deformed surface which is often referred to as a 'skin'. Therefore, the inside of the skin should provide some reference feature such as shadows [172] or markers, for example the TacTip [12, 13] (see Fig. 2.6). Alternatively, deformation of the skin can be inferred using stereo vision techniques [173], for example, the GelSight [174–176] (see Fig. 2.6) which provides detailed tactile images of 3D shapes and textures (see Fig. 2.10) and has recently been miniaturised for integration with robot hands and grippers [177].

A significant benefit of optical devices compared with all of the aforementioned technologies is their ability to naturally sense shear forces: skin movement perpendicular to the optical axis (tangential shear) is easily imaged by the camera. Consequently, however, these devices struggle with measuring entirely normal forces as the distance of the skin from the camera is more difficult to detect.

Camera based optical tactile sensors are generally cited as possessing high spatial resolution owing to the associated resolution of photosensitive arrays, although, in practice physical resolution of these devices also depends on the mechanics of the skin which can act as a low-pass filter.

These devices are traditionally thought of as bulky or cumbersome due to the requirement of containing an entire camera system, although recent improvements in affordable miniature cameras [178] and intelligent lensing techniques [177] has somewhat dispelled this label. Optical tactile sensors also require consideration of optics of the camera system; e.g., focal length, view angle etc. which present limitations on the shape and size of the entire unit or sensing surface.

The sample rate of such devices is defined by the frame-rate of the camera system and thus can be increased but often at the trade-off of added size, resolution or cost [179].

In contrast to the general approach of imaging a deformable skin with a camera, the PapillArray [180] is an example of a unique optical tactile sensor design. The sensor is designed specifically to measure static friction and detect incipient slip by using an innovative 3x3 array of pinhole cameras. Physical papillary constructed of individual silicone pillars contain a light source at the tip of the pillar and a pinhole aperture at the base. A 2x2 photo-diode associated with each pinhole is used to transduce the 3D force in the relevant pillar. The method employed by PapillArray is significant since it naturally enables both shear and normal force sensing.

Multi-modal sensors

As discussed in Section 2.1, tactile sensing encompasses the perception of a range of tactile dimensions and, indeed, the human fingertip is equipped with multiple tactile channels for the transduction of these dimensions. In response, an approach taken by some engineers is to develop multi-modal tactile sensors with multiple distinct technologies for the purpose of sensing a range of tactile dimensions. These devices commonly consist of an array based technology for encoding spatial properties of the stimulus and an additional low-resolution, high-frequency technology for

detecting vibrotactile or slip information.

A prominent multi-modal tactile sensor, known as the BioTac (see Fig. 2.6) contains an array of 19 sensing electrodes for high spatial resolution was developed by Wettels et al. (2008) [181]. The artificial fingertip is bordered by a flexible surface which is grounded. In between the sensing and grounded electrodes is a conductive fluid. External forces deform the fluid, resulting in a distributed pattern of impedance changes containing information about shape and force of contact [181]. The BioTac is also equipped with a single 'hydro-acoustic pressure sensor', which is essentially piezoresistive in nature [182]. This sensor directly measures the pressure within the conductive fluid and can be used for measuring high-frequency vibrations for texture discrimination perception [183]. Finally, a thermistor provides temperature sensation [7]. The entire sensor is roughly the size and shape of the human fingertip and is coated with an elastomeric skin with a fingerprint like structure [182].

Other less prominent multi-modal tactile sensors have been developed. For example, a multi-modal tactile skin [184] containing individual hexagon shaped pads each with a PCB equipped with multiple discrete sensors for temperature, acceleration, and proximity sensing. A multi-modal MEMS tactile skin is presented in [185] which can measure temperature with a nickel resistance device (RTD), thermal conductivity with a gold heater and nickel RTD pair and hardness using a strain-gauge. A compliant fingertip, containing a microelectromechanical magnetic, angular rate, and gravity system (9-DOF MEMS MARG), was used for detecting vibrations, and a deep MEMS pressure sensor, for detecting normal force, both embedded in silicone, is demonstrated in a surface categorisation experiment [186]. An optical tactile sensor [187] leverages an optical computer mouse chip to detect contact for high speed reflex and, in a separate mode, high resolution optical sensing was enabled at a lower frame-rate for accurate force estimation.

The weaknesses of multi-modal tactile sensors are generally associated with the core technologies used for individual modalities. For example, many of the aforementioned multi-modal sensors feature barometric MEMS devices which can't measure shear and tend to suffer from temperature drift. Additionally, there are associated drawbacks with including additional technologies to accomplish multi-modality in terms of added size, complexity and cost.

2.2.1.1 The TacTip

The research presented in this thesis is underpinned by previous work towards developing the TacTip (see Fig. 2.6). We leverage the TacTip's core sensing principles, manufacturing and some of the algorithmic work in terms of feature extraction and perception. Here we provide a brief overview of the history and key sensing principles to which we refer in subsequent chapters. More detailed description of the design and manufacturing processes are provided in subsequent chapters.

Chorley et al. [12] developed the TacTip in 2009, with an application to tactile edge sensing,

taking inspiration from the concept that morphology of epidermal ridges in the human fingertip is important for operation of Merkel cells (Fig. 2.7) (Section 2.1.1.1: SA-I) [39].

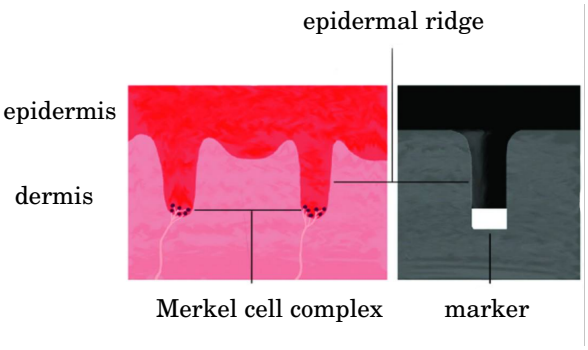

Figure 2.7: *Left*: epidermal ridges in the human skin. *Right*: corresponding pin in the TacTip. From [13].

The TacTip uses markers to provide an 'optical signature' of the tactile deformation of its skin [12, 13]. Although many variations exist, in its most generic form, the TacTip consists of a deformable 'skin' which is hemispherical in shape and filled with a soft transparent gel. Traditionally, the skin is constructed from a moulded black silicone rubber [188], however, more recently, it has been made using multi-material 3D-printing [189]. On the inside of the skin, an array of small 'pins' protrude inwards and, on the end of each pin, a contrasting white marker is either painted or 3D-printed. The pins are inspired by the hypothesis that contrastingly stiff epidermal ridges, in human touch, focus strain energy of skin towards the cite of Merkel cell complexes [39, 40]. Thus, markers are considered analogous to SA-I afferents. These markers are illuminated with a set of LEDs. Tactile information is transduced via lateral movement of the pins which are imaged using a standard USB camera. Examples of images captured by a TacTip camera are provided in Fig. 2.12.

Since the original study, the TacTip has gone through many design iterations [13] including integration with robot hands [25, 190] and novel shapes for the application to endoscopy [191]. A move from moulded to 3D-printed skin was notable for providing faster prototyping of new designs.

2.2.2 Encoding of Tactile Information

We believe that, as with human touch [23], it is useful to consider distinct tactile dimensions in robotics. We have seen that, depending on the tactile dimension being perceived, humans use different afferent types and employ a range of encoding mechanisms (Section 2.1.3). Here we aim to demonstrate that a distinction can be made between the same key tactile dimensions in robotics and, in-fact, it is useful to consider these tactile dimensions since it offers insight towards the best methods for transducing and encoding tactile information.

In analogy to human touch, trandsuction methods (sensor types) can be considered similar to afferent types in human touch. For example, a sensor that offers high spatial resolution with low sample rate; e.g., optical, could be viewed as analogous to the SA-I channel due to its high innervation density and preferential response to static stimulation. Once tactile data is collected, engineers generally perform some type of manual or automatic feature extraction from the transduced tactile data. We consider this process to be analogous to peripheral neural

encoding since in both cases the purpose is to robustly represent aspects of the data that are relevant to the tactile dimension being perceived in such a way that the percept can be easily extracted or decoded. In neuroscience a decoding model involves using neural features to generate a stimulus output [72]. In general, this description also aptly describes the process in robotics. More specifically, extracted features are mapped to a stimulus output via a model, usually a supervised machine learning algorithm; e.g., linear regression, Gaussian process or neural network.

Following, we review literature on robot touch relating to key tactile dimensions which mirror those in Section 2.1.3. We demonstrate that since the cues overlap with those in human touch, there are also many similarities in terms of transduction and encoding.

2.2.2.1 Weight

Perhaps surprisingly, the explicit perception of object weight has rarely been considered in robotics. A reason for this may be that, in practice, robots are seldom concerned with weight as a distinct dimension but rather, weight is an implicit object feature related to higher-level functional tasks such as object identification, grasp stabilisation and manipulation; i.e., an object's mass may be encoded in artificial tactile or proprioceptive signals (joint torques), (as is the case with humans: see Section 2.1.2), and in achieving higher-order functions robots may implicitly leverage this encoding. This is particularly the case for modern end-to-end techniques; e.g., Calandra et al. (2017) [4] who used cutaneous tactile information from the GelSight to predict grasp success with a CNN and later, Calandra et al. (2018) [192] where re-grasp policies were learned straight from visuo-tactile data. Further examples of end-to-end techniques for re-grasp are found in Li et al. (2014) [193] and Chebotar et al. (2016) [194] where the tactile data consists of readings from the BioTac's 19 impedance electrodes which are likely to implicitly encode object weight either intensively or spatially.

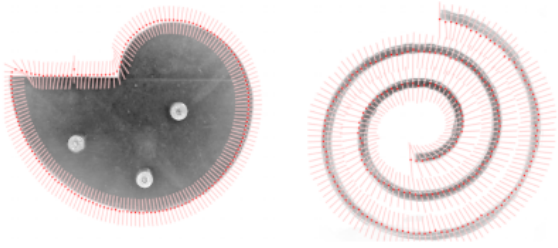

Figure 2.8: Examples of kinaesthetic shape reconstruction using the TacTip (from Lepora et al. (2019) [6]). Red points mark robot position and short pink lines depict the perceived edge orientation.

It is thought that human weight perception incorporates both cutaneous and kinaesthetic modes [76]. In robotics, it is common to use kinaesthetic or proprioceptive sensing for object recognition tasks which likely use object mass as an identifier. For example, Sinapov et al.(2011) [195] used joint torque sensor data of a robot arm whilst performing exploratory behaviours; e.g., shaking, lifting and dropping, to identify 50 common household objects.

2.2.2.2 Geometry

The definition of geometry for human touch, stated in Section 2.1.2, is also a pragmatic definition for robots. It is the curvature of an edge or surface and can be considered on two scales: small shapes that can be perceived with a single fingertip (cutaneous sensing) and larger shapes which require kinaesthetic inputs realised via haptic explorations such as edge tracing or grasping [23].

In robots, kinaesthetic geometry perception is generally manifested in closed-loop tactile control such as surface exploration or contour following using proprioceptive sensing and kinematics to trace the cutaneous sensor's location thus creating a map of the object's surface or edge. This concept is exemplified in an early study by Goldberg and Bajcsy (1984) [3]: a fingertip of resistive electrodes, controlled with a Cartesian robot, actively explores object surfaces and can thereby reconstruct object shape. More sophisticated control has been employed by Li et al. (2013) [5] and Martinez-Hernandez et al. (2013) [24] where edge orientation was perceived using array based sensors and leveraged as a process variable in PID-like control. Lepora et al. (2017,2019) [6, 189] used a CNN trained on TacTip images to regress over edge orientation and PID control to follow the edge of volute, circle and spiral shapes; resulting trajectories are shown in Fig. 2.8. Aquilina et al. (2019) [196] developed a method for kinaesthetically realising object geometries with the TacTip that works with continuous sliding rather than taps as seen with previous work [24, 189] and Driess et al. (2019) [197] demonstrate how low resolution force sensing can be used in an active exploration framework with multiple end-effectors to kinaesthetically attain object geometry in 3D.

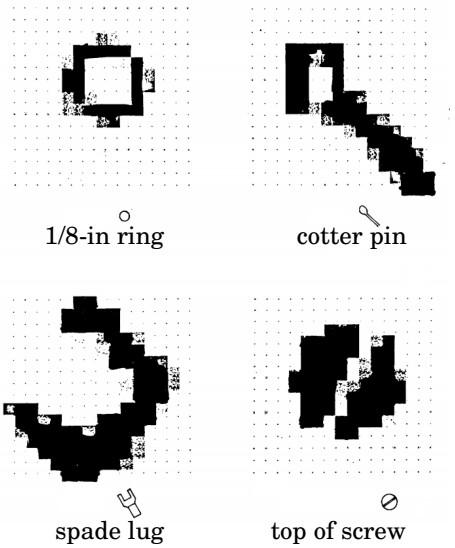

Figure 2.9: Tactile images collected with 16x16 resistive tactile array (from Hillis (1982) [2]).

In the cutaneous modality, it is thought that humans encode overall geometry spatially in SA-I response (Section 2.1.2). Similarly in robotics, array based sensors are typically used to produce 'tactile images' of the stimulus, which, by the spatial modulation of array response, naturally encode information about shape. An early example of this concept is provided by Hillis (1982) [2], who developed a resistive array of 256 taxels. Hillis did not perform any classification, although the resulting tactile images, shown in Fig. 2.9, demonstrate that this approach preserves much of the geometric information. Considering the spatial event plot (SEP) seen in Fig. 2.4 as an image transmitted to the CNS in populations of SA and FA afferent fibres, we regard the tactile images such as those shown in Fig. 2.9 as an artificial analogue. 38 years later, tactile images are still the standard method of transducing geometric information: such as those produced by the

GelSight (Fig. 2.10).

As with weight, cutaneous robot sensing of geometry is explicitly performed in few studies. Two examples of where geometry is explicitly sensed are provided by Roscow et al. (2016) [198] and Cramphorn et al. (2017) [199], where in both cases, edge sharpness was perceived using the TacTip. Marker positions create a tactile image from which sharpness was decoded with a likelihood model which treats each marker dimension independently, thus maintaining a spatial code. This, therefore, exhibits noteworthy similarities with theories of encoding of geometric information in humans, particularly the overall object shape under static conditions [91, 92].

Again, sharing a characteristic with weight, The geometry of an object is a key characteristic for higher-order tasks such as object identification, grasping and manipulation. Thus, it is implicitly sensed in many robot experiments. The trend in object recognition has been towards using image processing techniques to extract hand crafted features which encode spatial properties of the tactile imprint [172, 200–205].

The recent success of CNNs in image classification tasks has lead researchers in robot touch to consider these methods with tactile images. CNNs have a benefit over more traditional image processing methods for feature extraction in that they can automatically learn to identify suitable spatial features for the required task. In one of the first applications of CNNs to robot touch, Kwiatowski et al. (2017) [206] predicted grasp success on natural objects of a parallel jaw gripper equipped with two low-resolution (7x4) capacitive tactile arrays using a CNN trained on the raw labelled tactile images. The authors report an improvement of ~10% over a study using the same equipment with unsupervised feature extraction [207]. The GelSight is an obvious candidate for the application of CNNs to encode geometric features owing to the high resolution images it produces (Fig. 2.10). This is demonstrated by Calandra et al. (2017,2018) [4, 192] who predicted grasp success and effective re-grasp with a parallel gripper. Bauza et al. (2019) [208] has also shown that CNNs trained with GelSight images have the ability to map from tactile images to local shape geometries.

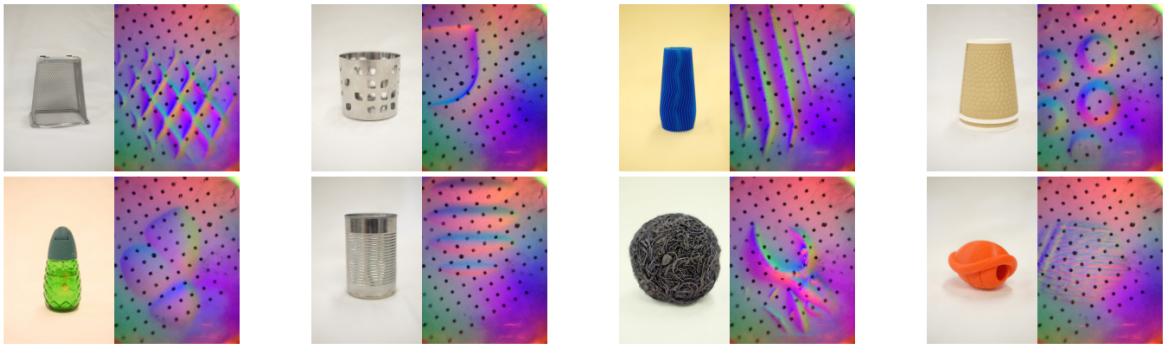

Figure 2.10: Tactile images collected with the GelSight (right) on natural objects (left) (from Calandra et al. (2018) [4])

2.2.2.3 Temperature

Thermal sensing is not often considered in robotics and is generally included only as an auxiliary modality in multi-modal systems. For example, Dahiya et al. (2011) [163] developed a multi-modal tactile sensing chip consisting of a 5x5 piezoelectric electric array and two temperature diodes, which, the authors claim, can be used to measure ambient as well as contact temperatures. Gupta et al. (2017) [209] present a flexible piezoelectric tactile sensor where the addition of barium titanate nano particles increased temperature sensitivity of the piezoelectric device. Gupta et al. demonstrated that capacitance changed with applied pressure and also contact temperature.

As mentioned in Section 2.1.2, humans employ thermal sensations to better identify objects [102]: materials such as metal have high thermal conductivity, thus feel cold, and therefore are easily discriminated from, for example, fabrics. Employing this concept, Xu et al. (2013) [7] used a single BioTac tactile sensor to classify 10 natural surfaces by fusing sensations from three distinct modalities. Temperature measurements from a thermistor embedded in the fingertip were used to determine thermal conductivity of each object. Combined with compliance and vibrational measures in an active exploratory framework, an impressive classification accuracy of 99% was achieved, however, a large contact time (15 s) was required to allow for heat flow.

2.2.2.4 Compliance

Compliance is an important perceptual dimension for robots: it is a salient characteristic for object identification and affects affordance for robot environment interaction. We have seen, in Section 2.1.2, that human compliance sensing of deformable objects involves a cutaneous modality thought to mediate information about shape of contact [83, 84]. The ability to actively explore the stimulus also permits a kinaesthetic mode which is believed to facilitate compliance through the relationship between displacement and applied force [86].

The principle of kinaesthetic compliance sensing was succinctly applied to robots by Shimizu et al. (2002) [144] who simply measured displacement of diaphragm pneumatically driven into compliant objects with a known force. Contrastingly, Drimus et al. (2011,2014) [145, 145] used only cutaneous data from an 8x8 piezoresistive array to identify fresh fruit from old, observing that contact area increased for old fruit due to increased compliance.

Both Su et al. (2012) [210] and Xu et al. (2013) [7] employ the BioTac for compliance sensing, leveraging both kinaesthetic and cutaneous modalities. Kinaesthetic modalities were realised by Su et al. by measuring applied force under fixed displacement and equivalently by Xu et al. through measuring displacement under constant load. In both cases, spatial properties of the output of the BioTac's impedance array were used to encode cutaneous information about compliance. Combined kinaesthetic and cutaneous compliance sensing is also demonstrated by Kappasov et al. (2018) [211] who distinguished natural rigid items from compliant versions of the same object; e.g., fresh vs. stale fruit. Kinaesthetic sensing was realised through closed-loop exploration by way of affordance, defined by force measurements from a tactile array. Spatial

properties of contact shape were encoded in the statistics of the tactile image; e.g., standard deviation.

An example of a modern tactile sensor, the GelSight, being used for compliance perception is seen in the work of Yuan et al. (2016) [212]. Hand-crafted features, derived from the tactile image, relating to change in contact area and normal force were identified and used in an optimised numerical model to estimate object hardness. The authors acknowledged that the identified features co-varied with object surface geometry. To address this issue, in a subsequent study, rather than hand-crafted predictors, Yuan et al. (2017) [8] leveraged a CNN to learn spatial features. Generalisation was improved, however, the model still showed a large drop in performance when tested on novel shapes.

In combining kinaesthetic and cutaneous sensing, the work of both Xu [7] and Kappasov [211] show clear similarities with human compliance sensing. Yuan et al. (2016,2017) [8, 212] omits any kinaesthetic sensing and notes poor generalisation across novel shapes. Considering that sensor deformation co-varies with hardness and object shape, it seems plausible that both cutaneous and kinaesthetic modalities are required in order to fully isolate the dimension of compliance. Indeed the work of Tan et al. (1992, 1993) [83, 84], showed that human participants had improved discrimination performance when permitted to actively explore objects thus enabling a kinaesthetic sense.

2.2.2.5 Orientation

As with both weight and geometry, perception of stimulus orientation is not usually considered as a stand-alone task in robotics but its use is frequently found in higher-order tasks. In particular, kinaesthetic reconstruction of object geometry often involves an intermediary step of inferring stimulus orientation relative to the sensor. It has been suggested that, in humans, tactile perception of orientation is primarily facilitated by a spatial code in primary SA-I afferent firing [43]. In robotics, we also observe that orientation is generally represented by spatial features extracted from tactile images which we consider analogous to spatial encoding.

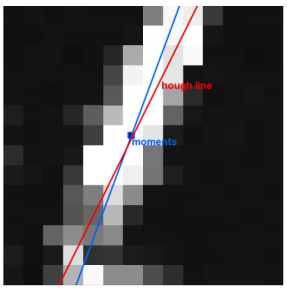

Figure 2.11: Orientation extraction using image moments (blue) from a tactile image collected with 16x16 piezoresistive taxel array (from Li et al. (2013) [5]).

For example, Li et al. (2013) [5] used image moments to extract orientation of an object edge from the tactile image produced with a 16x16 piezoresistive taxel array and the image moment was used as process variable in closed-loop control. An example of the tactile image and the feature extraction is displayed in Fig. 2.11. Kappasov et al. (2019) [213] extended this approach to extract orientation in 3D. Tactile images were gathered with 6x14 piezoresistive tactile array. Edge orientation about the sensor's z-axis was computed using PCA on the tactile image and the orthogonal orientation was

computed from the difference between centre-of-pressure and the centre-of-contact region. These orientations were used as process variables in a controller similar to the one used in [5].

The work in both [5] and [213] are examples of expert *feature engineering* since the encoding of orientation is so transparent that it can be directly related to the percept meaning control may be performed in feature-space. In contrast, most other approaches to using orientation in robot exploration involve first decoding the extracted features into the percept of stimulus edge orientation, thus performing control in percept-space; e.g., Martinez-Hernandez et al. (2013) [24] and Lepora et al. (2017) [189]. In the former, no manual feature extraction was performed: raw sensor readings from an iCub fingertip were used to perceive edge orientation using a probabilistic approach based on a histogram method of training data. In [189], TacTip marker positions were extracted using image processing and used to construct an a posteriori distribution over edge orientation classes using a similar measurement model as [24].

Lepora et al. (2019) [6] progressed the approach in [189] to use features automatically learned with a CNN. Examples of these raw tactile images, labelled with edge orientation, are shown in Fig. 2.12. In contrast to the tactile image shown in Fig. 2.11, the edge orientation is not self-evident, however, it is nonetheless encoded in these images, as is demonstrated by the ability to accurately following contour edges. Also with the TacTip, Aquilina et al. (2019) [196] exemplifies adept feature engineering: by using carefully selected shear invariant PCA features, the authors demonstrated that edge orientation could be accurately perceived even during sliding motions using Gaussian process regression, culminating in robust and accurate contour following using continuous sliding contact.

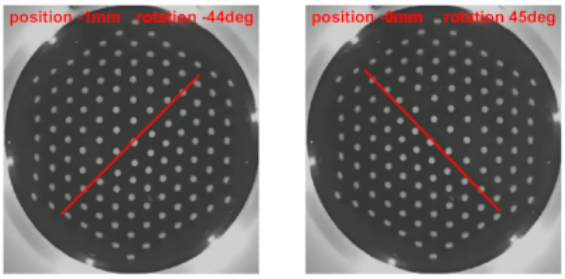

Figure 2.12: Examples of labelled TacTip tactile images from tapping on a disk edge (from Lepora et al. (2019) [6]).

As with [5] and [213], Li et al. (2014) [214] demonstrate how features extracted from high resolution tactile images can be directly related to object orientation. Feature descriptors were extracted from GelSight images and compared to a previously constructed tactile map to locate and orient a detailed planar surface (USB connector). By realizing the connectors pose relative to the robot, automatic robust insertion of the cable into a USB plug was performed.

The aforementioned literature on robotic orientation demonstrates that high-resolution tactile images naturally encode orientation in a spatial manner (as in humans) and that roboticists must consider high-level encoding that preserves this structure via either supervised feature learning [6], or extracting carefully considered hand-crafted features that either relate directly to orientation [5, 213] or map onto this dimension via a learned model [196, 215].

2.2.2.6 Texture

The distinction between coarse and fine texture perception is not as prominent in robotics as in human studies: the majority of robot experiments manifest as surface identification experiments; i.e., to discriminate or classify natural surfaces such as fabrics. These are less easy to categorise, but in general have microgeometry that might be considered fine; i.e., too small to be spatially resolved by human touch (<~1 mm) [42, 43, 109]. Therefore, rather than dividing robot studies according to stimulus scales, we choose instead to consider the transduction and encoding mechanisms. As with other tactile dimensions, the development of modern transduction techniques (sensors) has affected the available methods for perceiving texture and in recent years the progressions in deep learning has set the trend towards automatic feature detection.

Temporal Encoding of Texture

Temporal encoding of texture is the most common approach to robot texture perception. Generally, the chosen sensor provides low-dimensional, high frequency time-series data collected via sliding relative to a stimulus. From raw time-series data, engineers have leveraged many novel features to encode texture, typically using frequency components of the signal.

Mukaibo et al. (2005) [216] and Yi et al. (2017) [153] found that basic encodings based on peak frequency and simple statistical features derived from a spectral analysis of time-series data collected with strain gauges and a piezoelectric sensor, respectively, were good predictors of textures. These studies were however limited to simple artificial structures.

When testing with more complex structures, Edwards et al. (2008) [217] found that simple numerical features extracted from the frequency spectrum of audio data collected with a microphone sensor; e.g., modal power, were relatively poor predictors of texture compared with PCA features of the frequency components. Sinapov et al. (2011) [147] used even more complex stimuli in the form of 20 natural micro-textures; e.g., corduroy, leather and wood. Transient information was retained by using the acceleration phase which was encoded in a coarse spectral histogram (spectrogram), thus retaining both frequency and temporal structure. Good accuracy was achieved using an SVM to decode texture. Both of these studies indicate that harmonic structure might be important for discriminating stimuli that are more complex than artificial gratings. Similarly, Johnsson and Balkenius (2011) [218] found that spectrogram bins of tactile data collected on natural surfaces with a microphone were suitable features for clustering samples based on surface type.

The work of Ho et al. (2012) [219] appears to confirm that harmonic structure may be important for natural surfaces: probabilistic approaches and autoregression feature extractors applied to raw sensor data were ineffective for discriminating three natural surfaces (photo paper, denim and tape paper). Instead, an extensive set of statistical features (mean, variance, standard deviation, entropy and energy) were extracted from the discrete wavelet transform of each signal and an artificial neural network employed to decode these features achieved

classification accuracy of 90%. Although the features used by Ho et al. did not directly encode harmonic structure, information relating to the distribution is contained in statistics such as variance, standard deviation and entropy.

Fishel and Loeb (2012) [183] conducted an impressive study where 117 natural textures were classified with an accuracy of ~95% using cutaneous vibrotactile data collected from the BioTac's piezoresistive pressure sensor. In contrast to previous studies with natural texture, simple numerical features of the induced vibrations were employed. Each feature was directly related to a subjective percept of human texture perception: as proposed by Bensmaia and Hollins (2005) [220], signal power was understood to encode *roughness* and the spectral centroid was believed to relate to fundamental frequency therefore encoding *fineness*. The actuator current required to maintain consistent sliding speed was implied as a proxy for the subjective terms *stickiness* and *slipperyness* - dynamic friction. These perceptual dimensions have been shown to be orthogonal in human perception [221], which may explain the success of this technique in spite of the relatively simple feature set.

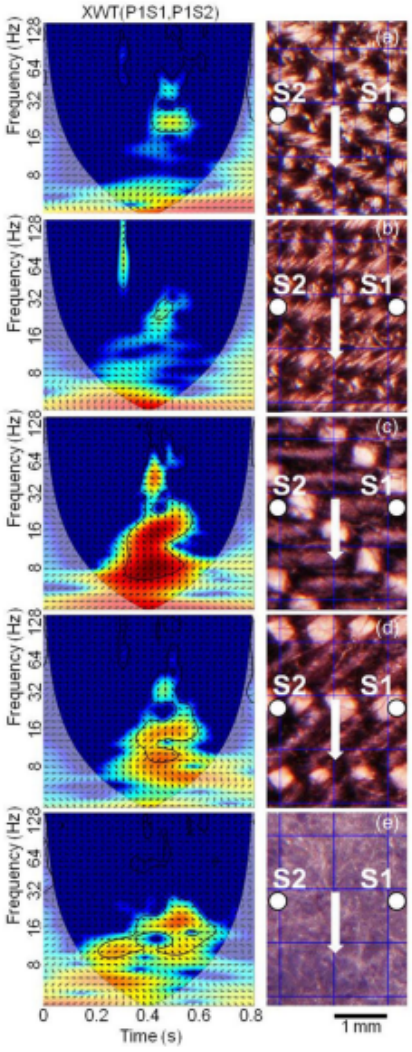

Figure 2.13: *Left*: cross-wavelet transforms during active rubbing of five textiles. Small arrows indicate relative phase of signal from sensors S1 and S2. *Right*: structure of each textile imaged by optical microscopy. The arrow represents rubbing direction [149].

In contrast to aforementioned literature, Dallaire et al. (2014) [222] used features extracted from time domain data. This, the authors suggest, encodes phase information which can allude to more complex structures of the stimulus. A MEMS accelerometer (800 Hz sample rate) detected vibrations when slid across natural textures at constant velocity and individual samples were constructed by segmenting time-series data into 1 s intervals. 7 numerical features were extracted; e.g., variance, skewness and sum of high frequency components, which were hypothesised as relating to vertical movement, irregularities and hardness respectively. An SVM trained on these features achieved an impressive classification accuracy of 99.95%.

Spatial Encoding of Texture

As described in Section 2.1.2, humans are thought to utilise spatial codes in SA-I and FA afferent response mediated by static touch for all but the finest textures. Examples of spatial encoding with static touch for texture

perception is observed in robotics, although, in general these studies are not put into the context of stimulus spatial scales and other modes of texture perception (duplex theory [57]).

The GelSight has often been applied to natural surface identification via static contact owing to the high resolution of tactile images. In fact, the spatial resolution of the GelSight far exceeds that of human tactile sensing and consequently the GelSight has been shown to perform well at identifying textures based on spatial codes where humans would likely require vibrational cues. For example, Li and Adelson (2013) [223] demonstrate classification of 24 natural fine-grained surfaces; e.g., wood, sandpaper and denim were classified to an accuracy of ~99%. 'Local binary patterns' (LBP), originally designed for image classification [224], were used to extract statistical features. A histogram of statistical features was constructed for each texture and classification was performed using the Hellinger distance as a similarity metric. Li et al. developed a method of encoding textural structure at different scales by extracting the same statistical feature after downsampling the image through a Gaussian pyramid, thus LBP regions of the same size would capture increasingly larger portions of the original tactile image. This technique is similar in purpose and principle to the max pooling operation in convolution neural networks.

As is the trend with other tactile dimensions, more recently CNNs have been leveraged to learn relevant spatial features. These techniques are particularly applicable to tactile images of texture collected with the GelSight. For example, Luo et al. (2018) [225] used the two parallel CNNs based on AlexNet [226] to learn features from tactile and vision data for the classification 100 natural fabrics. Outputs from the final dense layers were fused using maximum covariance analysis that maps the feature vectors to a shared latent space in which the covariance between them is maximised. Fusion results were ~90% vs. ~85% for vision or tactile alone.

Spatio-Temporal Encoding of Texture

Spatio-temporal methods utilise the interaction between spatial and temporal modulation of sensor response. These methods require sufficient spatial and temporal resolution of the sensor in order to capture the relevant information.

The work of Oddo et al. (2011) [149] demonstrates a spatio-temporal method whilst also cementing the idea that harmonic structure and/or temporal aspects of induced vibrations are required for more complex stimuli. Vibro-tactile data collected with a high frequency (300 Hz) low resolution (2x2) resistive array was shown to encode spatial period of artificial gratings in the frequency carrying maximum power. However, data collected on natural textures was less structured. The authors implement a complex encoding of cross-wavelet transforms (XWTs) (see Fig. 2.13) which provides a time-varying frequency spectrum as well as relative phase between pairs of taxels, thereby offering a spatio-temporal encoding. XWTs were qualitatively and consistently different between textures (see Fig. 2.13), suggesting that information contained in time-varying frequency spectrum could be employed to discriminate natural textures.

Yuan et al. (2018) [227] compared spatio-temporal with purely spatial methods. A single GelSight was mounted on one finger of a parallel jaw gripper and time series data were collected

during grasping with the aim to classify 11 textural properties of 153 clothing items. Properties included thickness, softness and durability and some semantic properties such as wash method. CNNs were used as spatial feature detectors. For the time-series model, the output of the CNN for each frame was input into an LSTM of 9 units. Only marginal improvements were observed when using time series data, indicating that very little salient information is contained within the temporal aspect of data collected during squeezing. This is perhaps not surprising considering that temporal aspects thought to be useful for texture perception in humans are related to induced vibration.

In an early study into the power of deep learning for texture perception, Baishya et al. (2016) [228] used 3D spatio-temporal data collected with flexible piezoresistive sensor from Tekscan of 4x4 taxels (750 Hz) during sliding, to classify 6 natural materials: metal, rough plastic, foam, paper, wood and smooth plastic. Their method learned spatio-temporal features via a slightly unconventional application of convolutional filters where the input data was first flattened in the spatial domain to produce 2D samples and asymmetric 2D filters were spatial vs. time in dimensionality. Good generalisation is observed with the CNN approach, with accuracies of 91% achieved on novel surfaces. The authors stated that structure in the spatial response of the tactile sensor might be less than that of temporal response justifying their application of convolutions to the time domain. This seems plausible given the spatial resolution of the sensor and the spatial scales of the natural textures used. An interesting comparison can be made to the work of Yuan et al. [227] where clearly the majority of salient information was contained in spatial aspects of the data. This difference is likely due to the resolution of the sensor data (high in [227] and low in [228]) and the type of tactile contact: where Yuan et al. used squeezing, Baishya et al. instead permitted vibrational cues via sliding contact.

An extensive comparison of pressing vs. sliding and hand-crafted vs. learned features for texture classification was performed by Taunyazov et al. (2019) [229]. Capacitive taxels on the iCub humanoid robot's forearm were used to acquire spatio-temporal data from pressing and sliding on 23 natural textures. Two types of hand crafted feature were employed: *(i) static features* consisted of standard deviation and mean across 60 taxels in a single frame and *(ii) dynamic features* were frequency based statistics as in [183], extracted only from the sliding touch data. Separate SVM models were trained with *static features* during presses and *dynamic features* during sliding, achieving accuracies of 61 % and 77 % respectively and 88 % in combination. Two SVM-LSTM models were trained with *static features* collected for each frame of the pressing and sliding data achieving accuracies of 61 % and 86 % respectively and 96 % in combination. These results suggest that a large portion of salient information is coded within simple statistics (*dynamic features*) based on vibrations induced with sliding, however, significant information is also contained in complex harmonic or phasic information which is unlikely to be retained in the dynamic feature set, however, may be decoded from the raw time signal used with the multi-layer LSTM model. Finally, two CNN-LSTM models were trained with raw data from the

touch and sliding modes, achieving accuracies of 85 % and 86 % respectively and, in combination, an accuracy of 98 % was obtained. Clearly only moderate improvements were observed when using learned spatial features attained with a CNN. It seems likely that this is a consequence of using a tactile sensor with low spatial resolution where most of the salient information is contained with temporal aspects of the data.

speed-invariance of Robotic Texture Perception

As stated, we are particularly interested in texture perception since it provides a broad representation of the entire tactile system in humans (Section 2.1.3). For fine textures, where dynamic touch is employed, we have seen that encoding enables speed-invariant perception [126] and therefore alludes to a scientifically important characteristic of the neural code. Thus, we consider this an important aspect to achieve in robot touch which models human sensing: if we are able to achieve speed-invariant texture perception it may provide insight into mechanisms of speed-invariance in human touch. Additionally, where robots are to replace humans, it is natural to endeavour towards the same capabilities.

An important aspect of human texture perception is its apparent invariance to scanning speed. To our knowledge, it has not been explicitly studied in robotics. However, the work of Romano and Kuchenbecker (2014) [230] alludes to this phenomenon. Vibrotactile data was collected by tracing a 3-axis accelerometer (10 kHz sample rate), held by a PR2 humanoid robot, across 15 natural textures whilst varying contact force and speed. The general approach was to encode frequency data in a coarse set of frequency bins and an SVM was used to decode texture class. The basic method yielded poor results which the authors attribute to novel test conditions resulting from varied speed and normal force. The authors combated this by providing speed and normal force as two extra features to the SVM which improved recognition rate to ~61%. The authors then proposed a non-uniform scaling of the frequency bins which modelled the logarithmic scaling in human perception [220]. Interestingly, this further improved performance by 3.4%.

2.2.3 Tactile Sensing for Robot Hands

Arguably, the primary function of tactile sensation in humans is for manipulation. Indeed, the highest density of discriminative afferents are believed to be located in the fingertip [42]. Thus, there is a natural inclination for roboticists to leverage artificial tactile sensing, particularly in the form of artificial fingertips when working with robot hands for high-order functional tasks such as object identification, grasping and manipulation. For an in-depth review of tactile sensing in robot hands, we refer the interested reader to a review paper by Kappassov et al. (2015) [1].

An early study by Kawasaki et al. (1999) [231] set a high benchmark for subsequent work towards endowing robot hands with a sense of touch. The authors presented a sophisticated, 16-DOF, anthropomorphic hand, the Gifu hand II. The hand was equipped with 624 resistive tactile pads distributed across all five fingers and the palm. The authors show varied tactile

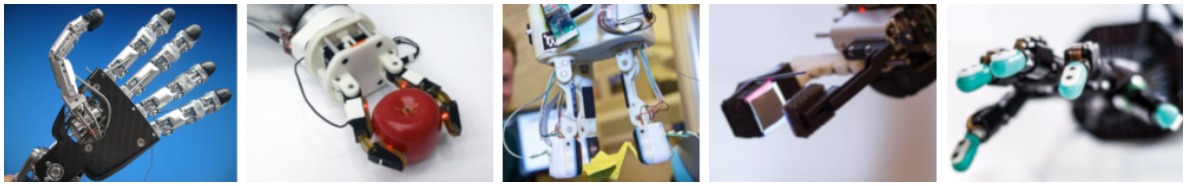

Figure 2.14: Examples of robot hands endowed with the modern tactile sensing technologies shown in Fig. 2.6. *Left* to *right*: iCub hand with iCub fingertips, OpenHand T42 underactuated gripper with TakkTile sensors, OpenHand Model O with TacTip sensors, Baxter hand with single GelSight sensor and the Shadow Robot hand with BioTac sensors.

profiles generated when grasping different shaped objects thus demonstrating that the objects are likely to be discriminable based on tactile information. The observed discernibility is owing to the high spatial resolution offered by the large number of tactile pads.

Since this influential study, a sense of touch has been used to aid many high-level tasks performed with robot hands, such as grasping [4, 192, 206, 207, 232], pose estimation [233, 234], object identification [140, 208, 235, 236], and manipulation [237, 238]. In grasping, there is a trend towards success prediction, slip detection, grasp stabilisation and effective re-grasp, all of which are significantly aided by, or arguably require, cutaneous tactile sensing. For the majority of grasping tasks, even in modern examples, it is sufficient to use simple two fingered, parallel jaw grippers, with tactile sensors on one or both of the fingers. For example, Kwiatowski et al. (2017) [206] used an under-actuated parallel jaw gripper from Robotiq, equipped with a 7x4 capacitive tactile array on each finger. A CNN, trained on raw tactile images was used to predict grasp success on natural, previously unseen, objects with an accuracy of ~90%. In the same year Calandra et al. (2017) [4] fused vision and cutaneous tactile information to predict grasp success using a Weiss WSG-50 parallel gripper, and two GelSight sensors, one for each finger. Also using a CNN, trained on raw tactile images, the authors achieved comparable accuracies on novel objects. We argue that, in both cases, grasp success signatures rely on spatial information, for example, where on the sensor the object is grasped, the contact area and the relationship between contacts of the two fingers. These features are likely to be captured by the array based capacitive sensor in [206] and the high-resolution GelSight in [4].

Slip detection and friction estimation in robotics is extensively reviewed by Chen et al. (2018) [239] in terms of both its fundamental principles and application with robotic grippers. In particular, the authors argue that slip detection and friction estimation are essential for stable grasping and manipulation. There is some contention surrounding the nature of slip and 'incipient slip' (a phenomenon believed to precede the actual slip event by fractions of a second [240]), with some researchers subscribing to the idea that vibrotactile features are key for effective slip detection [241, 242], thus requiring sensors with high sample rates, whilst others observe that lower frequency sensors can achieve detection of object slippage via spatial signatures [243], therefore requiring higher resolution tactile sensors, or simple changes in tangential load [244].

Object identification with robot hands is naturally performed by combining cutaneous and proprioceptive data. For example, Schmitz et al.(2014) [140] leveraged joint angles, high resolution cutaneous sensing and 6 axis force torque sensors in each finger of the TWENDY-ONE hand to classify natural objects in one of the first applications of deep learning to tactile sensing. Spiers et al. (2016) [236] used an adaptive, under-actuated 2-fingered gripper equipped with 16 low resolution TakkTile sensors, where tactile and proprioceptive data were fused to identify grasped objects using random forests. In contrast, Vásquez et al. (2016) [235] used a simulated 5-fingered anthropomorphic hand to classify generic shapes using only proprioceptive data and, Bauza et al. (2018) [208] used cutaneous-only sensing of tactile images from the GelSight to perform surface reconstruction from multiple grasps with a parallel jaw gripper using CNNs.

In-hand manipulation could be considered one of the final frontiers of robotics. There have been attempts at demonstrating robot dexterous in-hand manipulation which make use of touch. For example, in [237] where tactile sensors were used with the Shadow 5-fingered anthropomorphic hand to translate on object to a desired pose, and in [25], where a two fingered robot hand equipped with a single TacTip style sensor was used for in-hand rolling. These demonstrations of in-hand robot manipulation have been somewhat surpassed recently by the accomplishments of OpenAI, described in a manuscript by Andrychowicz et al. (2018) [245]. The authors used deep reinforcement learning, trained with vast amounts of real and simulated data to demonstrate high-levels of dexterity when manipulating a Rubik's cube with a single Shadow anthropomorphic hand. Tactile sensing was omitted from their research since it is notoriously difficult to simulate. Instead the authors used only vision and proprioceptive information. The shear quantity of information required in [245] demonstrates the significance of the challenge of in-hand manipulation and the fact that touch was omitted demonstrates the potential significant improvements still: as argued by Chen et al. (2018) [239], the ability to estimate or simulate friction has the potential to vastly improve performance and/or training. However, researchers have a long way to go before touch sensing can be seamlessly integrated with such systems owing to the problem of its simulation. It is likely that, in the future, more sophisticated algorithms, perhaps derived from biological systems, will be more data efficient thus negating the requirement for simulation and enabling researchers to more easily leverage the modality of touch.

2.3 Conclusions

Robot touch is considered an important technology, as is evident from the quantity of research within the field. It also clearly presents many challenges as is demonstrated by the variety of approaches taken both algorithmically and in terms of hardware.

Our philosophy, as with many other researchers, is that human touch can provide valuable guidance of how to engineer artificial systems to suit the task at hand. In particular, we have presented human and robot touch in parallel, drawing analogies of sub-modalities with trans-

duction methods, and peripheral neural codes with feature extraction. We propose that through considering specifically these similarities, the development of artificial touch systems may develop faster and see increasingly practical application.

CHAPTER 3

HIGH-RESOLUTION AND HIGH-SPEED DUAL-MODAL OPTICAL TACTILE CONTROL

In this chapter we present a novel dual-modal optical tactile sensor which is a development of the TacTip. The TacTip's high resolution sensing is adapted to work with a low resolution camera system and an additional high-speed modality is enabled, capable of operating at significantly higher frequencies.

The work presented in this chapter has been published in the following peer-reviewed publication (a contribution statement is included):

- Pestell, Nicholas, John Lloyd, Jonathan Rossiter and Nathan F. Lepora. "Dual-modal tactile perception and exploration." IEEE Robotics and Automation Letters 3.2 (2018): 1033-1040.

  Contribution statement: N. Pestell realised the concept and wrote the manuscript with advice from J. Rossiter and N. F. Lepora. N. Pestell developed the dual-modal tactile sensor hardware and software and performed experiments. N. F. Lepora developed the biomimetic perceptual algorithm for high-resolution sensing and J. Lloyd developed the robot code, both of which were adapted by N. Pestell to work with the novel dual-modal tactile sensor.

3.1 Background

The TacTip is heavily based on the morphology and function of Merkel cells [12, 13]. In particular, marker positions provide a high-resolution, non-adapting signal which is ideal for perception of highly-detailed, static stimuli via spatial mechanisms, such as edge location and orientation [6, 188].

As described in section 2.1.1.1, the human tactile system is comprised of multiple channels to enable the perception of all key tactile dimensions (Section 2.1.3). Broadly speaking, the TacTip lacks the capacity for sensing with high-speed which is useful for responding to dynamic stimuli. To this end, here we aim to endow the TacTip with an additional 'high-speed' modality which we argue provides benefits for low threshold contact detection and reflex action.

We have designed and built a tactile sensor using a small image tracking chip (ADNS-3080 [246]) as an image acquisition system which offers both a high (temporal) speed (single value at ~2000 Hz) and high (spatial) resolution sensing mode (30x30 pixel image, ~3 Hz). The manufacturing cost of this sensor is low (~£100) and the body is completely 3D-printed, making the manufacturing process relatively simple. We also implement a novel approach to high-resolution sensing based on using raw pixel values as features without the need for any image processing, distinct from past work with 3D-printed optical tactile sensors [12, 25, 26, 189, 247, 248]. We draw the distinction between **multi-modality** and **dual-modality**, where the former defines sensing where more than one mode can be used at the same time. In contrast, the sensor presented here offers dual-modality which means that each mode must be used in isolation.

Contact detection is an important aspect of the human somatosensory system. It is crucial for manipulation and many higher order control processes. By comparing psychophysical and neurophysiological thresholds, Johnansson et al. (1979) [51] demonstrated that FA and/or PC afferents are responsible for low threshold contact detection and were able to rule out both SA-I and SA-II afferents for this purpose. Interestingly, the authors suggest that, if FA afferents are responsible for contact detection, contact is signaled by a single impulse produced by one or very few FA afferents [51]. This demonstrates the extremely low spatial resolution of the contact detection signal.

An additional artificial tactile modality that is able to quickly detect low threshold contact forces, similar to FA and/or PC afferents may be a useful aspect of robot touch. Based on the hypothesised biological mechanism for contact detection [51] an artificial analogue is not required to posses high-spatial resolution and, given a fixed bandwith, can therefore take advantage of increased sample-rate. Indeed, it is known that FA and PC afferents are responsive to significantly higher-frequnecies than either SA-I or SA-II afferents [44, 49]. We therefore argue that the engineered high-speed modality offers characteristics and function that are similar to FA and PC afferents.

As well as contact detection we also draw the loose analogy of the high-speed modality to that of reflex action in nature. In reflex action the tactile signal bypasses the brain and directly informs motor control to avoid potentially dangerous situations in as short a time as possible [31]. These types of stimuli are detected by free nerve endings and are quickly transported along myelinated Aδ fibres [9]. Because the high-speed mode offers faster sensing at a low spatial resolution it may offer benefits towards fast motor control for avoiding potentially damaging situations.

This kind of approach is taken by Knoop and Rossiter (2013) [187]. Here, a dual-mode compliant optical tactile sensor estimates force with two distinct modes: *(i)* a high-resolution 'explore mode', where image processing is used to infer motion of the sensor skin from a sequence of images from which force is accurately estimated and *(ii)* a high-frequency 'reflex mode' is used for quick reactions and rough estimation of force. A single optical device is used for acquiring images which can return data with high frequency at a cost of resolution or vice versa.

The performance of the presented sensor is tested in a number of tasks aimed at understanding the capacity for the developed high-speed mode for depth modulation which we argue is a proxy task for reflex action and, additionally, the high-resolution mode for edge perception. Finally, we combine the use of the two modalities into a single task of following the contour of a previously unseen object using the high-speed mode to first detect contact of the object, similar to a hypothesised function of the natural FA and/or PC channel [51], modulate to the correct depth and locate the edge of the object. Subsequently the robot switches to the high-resolution mode for perceiving edge angle and radial position. These percepts are then used in a control framework for contour following.

In [187] the sensor is used purely as a force estimator, whereas, in the presented study we use both modalities to perceive multiple dimensions (angle, radial position and depth) and use the perceived state to control the motion of a robot in a closed loop exploration task.

3.2 Development of a Dual-Modal Tactile Sensor

The presented sensor, shown in Fig. 3.2, is an evolution of the TacTip [**?**]. We adopt the core bio-inspired approach to transduction (Section 2.2.1.1) and many of the fabrication methods from a series of publications with the TacTip [249]. We aim to address the need for cheap tactile sensing with multiple modalities by replacing the CCD webcam used in the TacTip with a tracking image sensor similar to that used in [187]. We take a completely novel approach to extracting information from the image, where no-image preprocessing is used to reduce dimensionality: raw pixel values from a 30x30 pixel image are used as features for classification of tactile stimulation.

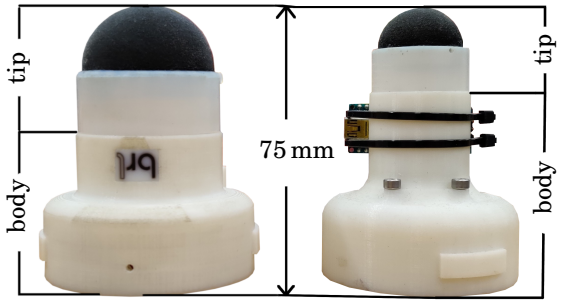

Figure 3.1: Side-by-side comparison of the TacTip (left) and the developed dual-modal tactile sensor (right).

3.2.1 Dual-Modal Sensor Design

The presented sensor, shown in Fig. 3.1 (right), is comprised of two main components: a compliant tip and a rigid body, housing the optical sensing element and electronics. The two components interlock via a bayonet fitting.

The tip is 3D-printed and consists of an outer 'skin' (Tango Black+ (Shore A 26-28)) and a rigid bayonet fitting (Vero White). Both elements are printed as a single part. The skin is deformable, enabling transduction of tactile information; it has the added feature of making the sensor compliant which is important when interacting with delicate or unknown objects. The inside surface of the skin (Fig. 3.4) features a concentric pattern of white dots (Vero White) on the end of short pins (Tango Black+). Once printed, the space between the inside of the skin and an inserted acrylic lens is manually filled with a clear silicone gel (RTV27905, Techsil UK (~Shore OO 10)). The gel provides stiffness to the tip which helps to minimise hysteresis whilst still enabling compliance. The overall diameter of the tip is ~27 mm which is a ~33% reduction from previous TacTip versions [12, 247].

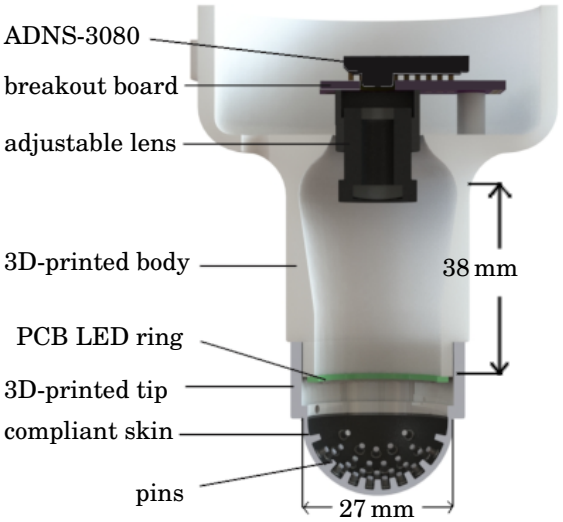

Figure 3.2: Computer modelled cross-section view of the sensor assembly. The two main components are *(i)* the 3D-printed body, housing the ADNS-3080 image tracking system and the PCB LED ring and *(ii)* the 3D-printed tip with a compliant sensing surface and pins.

The sensor body (ABS thermoplastic) is 3D-printed. It is responsible for housing the image tracking system. The selected shaft length (~38 mm) and diameter (~23 mm) were chosen in order to enable a full view of markers whilst minimising the shaft length which was underpinned by the focal length and view angle of the image tracking system. It also features a bayonet mount for easy replacement of modular tips and houses a PCB ring of six LEDs for illuminating the inside of the tip.

3.2.1.1 Camera System

The principal novelty of the presented sensor is the optical sensing element. Where the TacTip features a standard camera system, we use a tracking image sensor, *ADNS-3080, Avago Technologies* [246], designed for use in optical computer mice. The sensor combines an image acquisition system (IAS) which uses Complementary Metal-Oxide-Semiconductor (CMOS) technology to

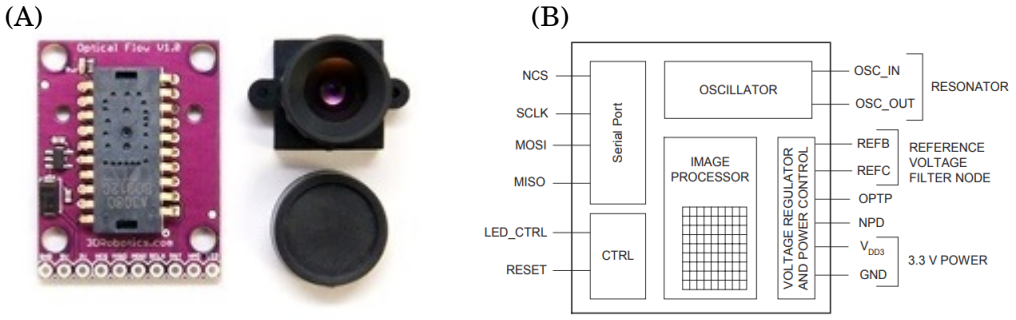

Figure 3.3: **(A)** Photograph of the ADNS-3080 chip mounted on a breakout-board (left) and adjustable lens mount and lens cap which fixes to the reverse side. **(B)** A block schematic diagram of the ADNS-3080 chip, showing I/O pins and image processor (IAS and DSP) [246].

acquire a 30x30 pixel image, and a digital signal processor (DSP), which processes the images to generate a series of statistics into a single chip [250].

A DSP is a specialised microprocessor. They are designed for a bespoke task, usually to carry out fast arithmetic operations with low power consumption, unlike typical microprocessors which are intended for more general applications. The speed of DSP operations are achieved by using an instruction set which is considerably smaller than those used for general microprocessors; e.g., X86 and ARM, and is optimised for performing the intended arithmetic operations. A common function of a DSP is to process, in real-time, a digital signal (e.g., video or audio) post analogue to digital conversion. Typical processing operations include fast Fourier transform, finite impulse response (FIR) filtering and convolutions. Due to their low power consumption DSPs often find applications in mobile technologies such as mobile phones [251].

The frequency of the IAS has a default of 2 kHz and is programmable to a max value ~6.4 kHz. Importantly, the sensor has two modes of operation: high-speed statistic mode, where a single value is available at the IAS frequency; and frame capture mode, where the sensor sends each pixel value of the 30x30 pixel image over the serial interface in sequence, thus the frame rate which is ~900 times lower than the high-speed statistic mode. Details of how we leverage these two modes to enable multi-modal tactile sensing are provided in Sections 3.4 and 3.5.

The presented sensor uses an ADNS-3080 on a breakout-board with an adjustable lens (focal length = 4.2 mm, view angle = 86°). A photograph of the ADNS-3080 chip and breakout-board is shown in Fig. 3.3 and a block diagram of the ADNS-3080 chip components is shown in Fig. 3.3.

3.2.1.2 Interfacing

To retrieve data from the ADNS-3080, it is interfaced with an 8-bit 16 MHz microcontroller, *Arduino Nano*, via a synchronous serial port connection (serial peripheral interface, SPI). For the purpose of conceptualising hardware interfaces, the tactile sensor is considered to be simply comprised of the ADNS-3080 and the Arduino Nano (see Fig. A.4; Appendix A). We interface the

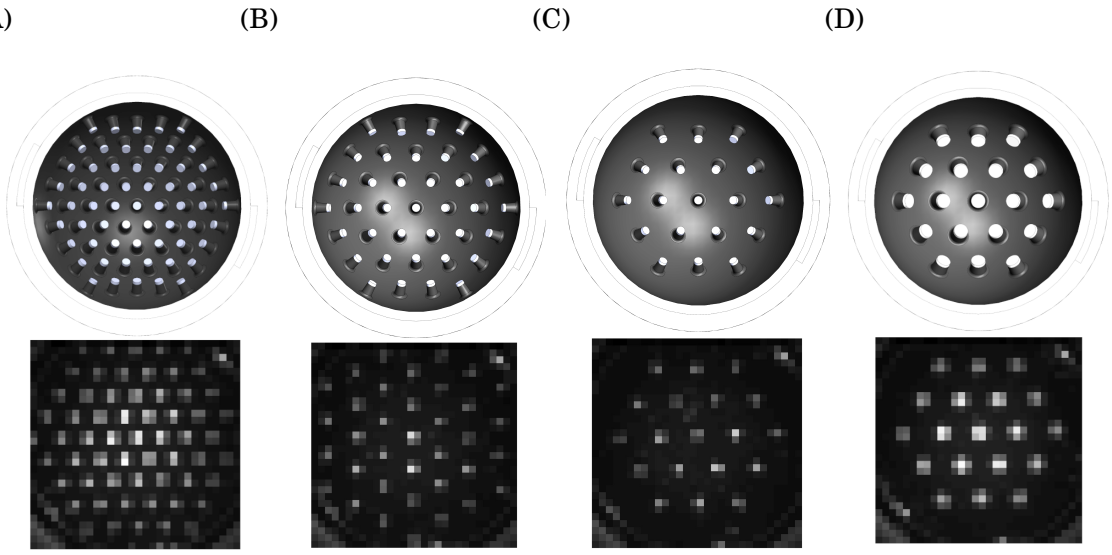

Figure 3.4: Computer models of four alternate pin layouts (top row) and respective tip images captured by the ADNS-3080 below. Tip-A features the same pin size and spacing as the latest TacTip version - diameter ~1 mm, spacing ~3 mm; Tip-B - diameter ~1 mm, spacing ~4 mm; Tip-C - diameter ~1 mm; spacing ~4.5 mm; Tip-D - diameter ~1.7 mm, spacing ~4.5 mm.

tactile sensor with an external PC through an asynchronous serial transmission via the Arduino's serial port and a virtual COM port on the PC.

For a detailed explanation of the interfacing procedure, including hardware diagrams and communication protocols, please refer to Section A.1.1 of Appendix A.

3.2.2 Alternate Pin Layouts

In previous work, the pin layout used with TacTip has proven effective for discrimination tasks [248]. However, in this study, rather than use pin positions, we apply the novel approach of using raw pixel values as tactile features and consider that the pin arrangement used in previous work may not be so well suited to this technique. When using pin positions each pin is tracked from rest to its position under stimulation, this means that each varying stimulation is likely to map to a very unique set of pin positions. This is not the case when using pixel values: each pixel remains fixed and therefore its value may have a more ambiguous relationship to the nature of the stimulation. For example, drastically differing stimulation could result in similar values for one, or a number of pixels, if different pins are imaged by the same point on the CMOS array, thus the data is ambiguous. We therefore hypothesise that a less dense pin layout will result in fewer pin overlaps and a possible improvement of tactile perception, however, this may come at a trade-off with spatial resolution: it is likely that more markers enables a better capacity to transduce fine spatial detail.

Four alternate pin layouts are considered, shown in Fig. 3.4. Tip-A features a pin spacing of

~3 mm and pin diameter of ~1 mm, these dimensions were chosen to match the latest TacTip version. Tips-B and C have the same pin diameter as the latest TacTip version but with increased pin spacing of ~4 and ~4.5 mm respectively. Finally, Tip-D has a pin spacing of ~4.5 mm but with an enlarged pin diameter of ~1.7 mm. The patterns are arranged as projections from the equator of the sensing hemisphere onto the inside surface of each tip. In this way the pin spacings are even as viewed from the camera. The bottom row of Fig. 3.4 shows each tip as imaged by the ADNS-3080.

3.3 Robotic System and Software Structure

A high-level block diagram of the overall software structure is shown in Fig. 3.5.

We communicate with the sensor via a virtual COM port on the PC (Section A.1.1.2; Appendix A) with a Python class, set up as a sensor-client, which implements the PySerial library [252]. This class contains methods for triggering read-write operations on the Arduino-ADNS system (Fig. A.4; Appendix A). A server script runs on the Arduino, in a continuous loop, which checks for incoming read-write triggers on the serial port and then executes the required read-write operation with the ADNS-3080 by implementing the Arduino SPI library. Arduino code for the dual-modal sensor has been open sourced and is available on GitHub [253]. The relevant tactile data is then sent back to the main PC via the serial port.

We enable the sensor to interact with its environment with the use of a six degree-of-freedom robot arm (IRB 120, ABB Robotics) to which our sensor is mounted as an end-effector. The arm can precisely and repeatedly position the sensor with an absolute repeatability of 0.01 mm. A custom 3D-printed mount is bolted to the rotating (wrist) section of the arm to which the sensor is attached via a bayonet mechanism.

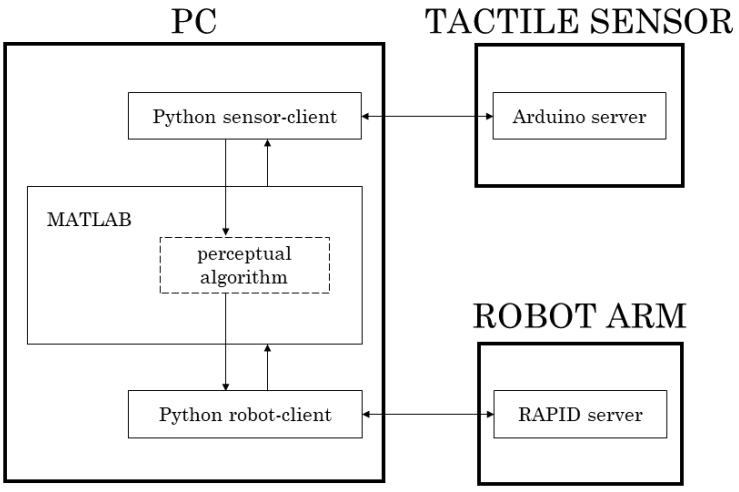

Figure 3.5: Diagram of high-level software structure for the robotic system.

For controlling the arm we leverage the open_abb open source software project [254]. open_abb provides a Python client API in the form of a robot class from which we can instantiate a robot object and call methods which perform operations such as Cartesian moves, joint moves, setting the speed etc. open_abb also provides a server script written in RAPID, the ABB controller native language. This runs on the ABB control unit in a continuous loop. When we call methods on our robot object, commands are sent, over a network, to the RAPID server via TCP and placed on a software buffer. The RAPID script will process the data within this buffer and command the robot accordingly.

High-level programming for different routines and tasks is performed in MATLAB: Python client methods for recording from the sensor and controlling the robot arm are called from a MATLAB experiment script. In this way, sensing and motor control are synchronised and closed-loop control policies can be implemented. Between retrieving data and implementing a control strategy an intermediate tactile perception algorithm is used to map sensor observations onto relevant state variables (e.g., position, orientation). This is also written in MATLAB. Details of the perception are provided in Sections 3.4.1.1 and 3.5.1 for high-speed and high-resolution modes respectively.

3.4 High-Speed Sensing

The ADNS-3080 is used in high-speed mode to enable the first distinct modality (mode-HS). The illuminated dots on the inside of the sensing surface provide visual features which the IAS detects. The DSP uses sequential frames to generate statistics based on the motion of these features when the sensing surface experiences stimulation. The statistics are: motion detection, relative x and y movement, shutter opening and closing times, maximum pixel value and surface quality (SQUAL) and are described in Table 3.1.

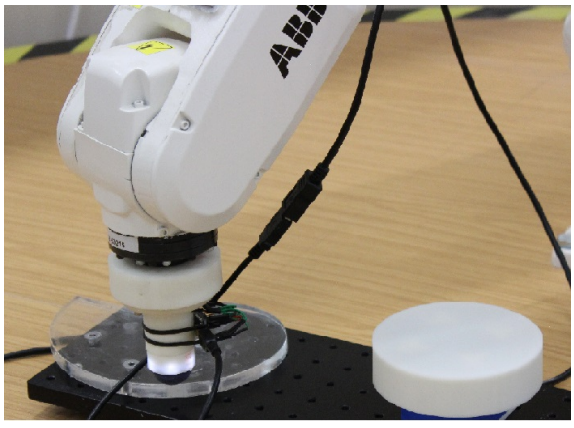

Figure 3.6: Sensor mounted as an end-effector on the robot arm used for experiments. The two stimuli used are also shown: circle (right) and non-uniform volute (left).

To retrieve statistics, the ADNS-3080 is sampled in 'burst mode'. Burst mode enables a faster rate of data transfer when sending or receiving data from multiple registers by continuous data clocking from a sequence of registers rather than addressing each register in order. A detailed description of burst mode operation is provided in Section A.1.2 of Appendix A.

Table 3.1: Description of the six image statistics from the ADNS-3080 available in burst mode [246].

	Description	Register
Motion	Motion since last report or PD, MSB = 0 - No motion, MSB = 1 Motion occurred, data ready for reading in Delta_X and Delta_Y registers.	0x02
Delta_X	X movement is counts since last report. Absolute value is determined by resolution. Reading clears the register.	0x03
Delta_Y	Y movement is counts since last report. Absolute value is determined by resolution. Reading clears the register.	0x04
SQUAL	Surface quality is a measure of the number of features present in the image. Surface quality is typically maximized when the imaged surface is at the optimum distance from the imaging lens (the nominal Z-height).	0x05
Shutter_Upper **Shutter_Lower**	Units are clock cycles. Read Shutter_Upper first, then Shutter_Lower. They should be read consecutively.	0x0e 0x0f
Maximum_Pixel	Maximum Pixel value in current frame. Minimum value = 0, maximum value = 63. The maximum pixel value can vary with every frame.	0x07

Of course, one could attain similar statistics on image sequence data using a PC, however, by leveraging the on-board processing provided by the ADNS-3080's DSP component, we are able to gain tactile data at a considerably higher frequency because relatively few bytes of data are transferred per frame compared with sending a whole image. We argue that the low-resolution high-speed nature of mode-HS will provide benefits for fast low threshold contact detection in a similar fashion to a hypothesised function of FA and/or PC afferents [51]. We also argue that both of these aspects may have benefits for providing an analogue of natural reflex action [31] in robotics, e.g, for providing a fast signal when normal pressure has exceeds a threshold value that could potentially damage the sensor.

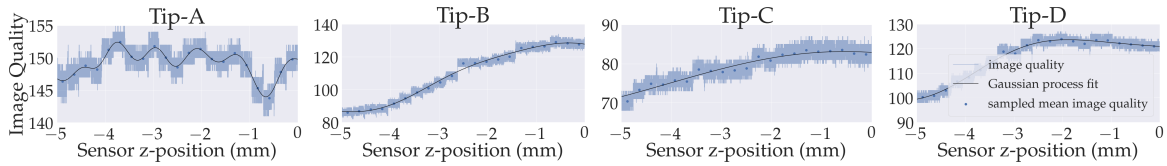

Figure 3.7: Surface quality statistic plotted against depth of compression on a flat stimulus for the four pin layouts shown in Fig. 3.4. The raw data (shown in grey) was recorded at constant depths with a 0.25 mm separation for 5 seconds at each depth. The red markers show the mean value at each depth and the line is a Gaussian regression fit.

3.4.1 Task: Contact Depth Control

Here we consider the task of controlling the contact depth, z: we require the robot to follow a target trajectory where position is modulated by measurements from the high-speed mode of the tactile sensor.

One of the most common ways in which the TacTip becomes damaged is over exertion of normal force which causes the tip to rupture. We therefore argue that the contact depth control task is a valid proxy for reflex action since it demonstrates the ability to sense and respond to potentially adverse conditions. To avoid damage it is important that the overall system performs with as little latency as possible. The speed of sensing is an important aspect of this system and therefore we intend to examine the capabilities of our high-speed sensing methods in this area.

3.4.1.1 Data Collection and Calibration

For the four alternate pin layouts shown in Fig. 3.4, image statistics were collected for 5 second intervals at constant compressions on a flat stimulus. $N_z = 20$ depths were used spanning -5 mm $\leq z_i \leq$ 0 mm. Measurements are taken relative to the where the sensor tip makes initial contact with the stimulus. An average value at each step is then used to fit a Gaussian process regression (MATLAB, fitrgp) which is used as calibration in the depth control task.

Of the 5 statistics, only surface quality provided a useful response. All other measures were unresponsive when exposed to tactile contact which is consistent with observations made previously [187]. Fig. 3.7 shows the surface quality (blue) plotted against contact depth for the four considered pin layouts. Also displayed is the average value for surface quality at each step (blue markers) and a Gaussian process regression (black line), fitted to the average values. The sensor exhibits a highly non-linear relationship between surface quality and contact depth with Tip-A. Tips B-D exhibit a general relationship of decreasing surface quality with increasing contact depth. We note that only Tip-D does not show a monotonic relationship and between Tip-B and Tip-B, Tip-B shows the largest range in measured surface quality within the sensor depths used.

Pin layout clearly has a large influence on the high-speed sensing modality. Surface quality is a measure of the number of features present in the frame [246]. It appears to be maximised when pins are at the optimum distance from the lens and reduces as the pins move out of focus. We suggest that Tip-B has the best balance between pin spacing, number of pins and pin size for the surface quality algorithm to detect each pin as a separate feature. The ADNS-3080 may struggle to separate the pins of Tip-A as separate features, whilst tips C and D have fewer pins, yielding a lower overall value in surface quality.

3.4.1.2 Control

To test the depth perception, a task is executed where the robot is required to follow a sinusoidal target trajectory in z whilst remaining in constant contact with the flat stimulus used for calibration. The control is a closed-loop system using only tactile feedback from mode-HS to modulate the z-position of the sensor.

Before each step move in z, one sample from the high-speed mode is taken, after which we estimate the depth according to the calibration curve. The robot then modulates its z-position using a control policy, π, by performing a relative move from the current perceived depth towards a target, which is defined by sinusoidal trajectory in z,

$$\Delta z_i = \pi_i[z_{\text{dec}}] = (z_{\text{target},i} - z_{\text{dec}}), \tag{3.1}$$

z_{dec} is the perceived depth and $z_{\text{target},i}$ is the target at step i. Here we use just one sample for predicting the depth in order to minimise the reaction time and benefit from the high sampling rate provided by mode-HS.

3.4.1.3 Results

We used Tip-B and the surface quality statistic for this test since this combination was found to produce the best predictor of depth (Fig. 3.7). Results are shown in Fig. 3.8.

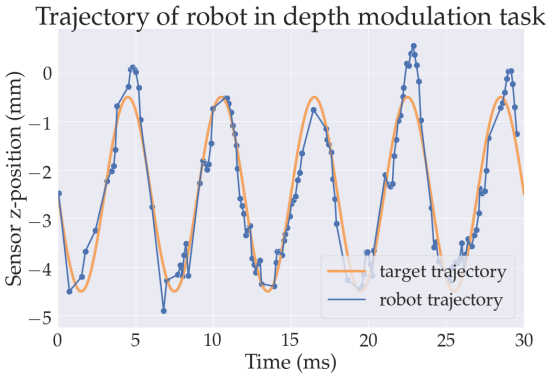

Figure 3.8: Sensor trajectory in z (blue) and target (red) vs. time. Blue markers show the sensor position at each step.

The sensor successfully follows the target to within ~1.1 mm and a mean absolute discrepancy of 0.37 mm. Divergence from the target is mainly observed at the maximum z-position which may be explained by hysteresis: magnified by a fast step rate, the compliant tip requires time to equilibrate after each step. Also, sensor noise (see Fig. 3.7) may have a negative impact on accuracy of the prediction. The effect of noise would be reduced by using the mean of multiple samples at each step, as with training. This, however, would compromise sensing speed which is the significant quality of mode-HS. It is worth noting that the step-rate here is defined, not only by the sensor bandwidth but, more directly, by latency in the robot control system (which, as can be observed from Fig. 3.8, is inconsistent): a finite amount of time is required for the client to send control commands over a network to the robot server and for the robot to execute that command. With a more sophisticated control system, including a differential term in the controller to avoid overshoot, the step-rate and the frequency of the target trajectory could be increased although this was not a focus of the present work and

we stress that the use of mode-HS minimises the response time and a lower speed sensor would reduce the step rate further.

The surface quality statistic is dependant on the shape of deformation and therefore the stimulus; i.e., the z-position relative to the top of a stimulus does not map to surface quality in the same way for all stimuli. Therefore, it is most likely that the method would need to be re-trained if using stimuli other than a flat surface.

3.5 High-Resolution Sensing

In this modality (mode-HR), we make use of the whole 30x30 pixel image (see Fig. 3.4). A higher resolution of tactile sensing, compared with mode-HS, is attainable because of the increased dimensionality. This comes at a cost of sample rate (~3 Hz), which is significantly lower than mode-HS because each pixel, represented by a single byte, is sequentially piped over serial interface to the Arduino.

White markers, located on the tip of each pin, provide optical output for transduction of tactile information, whilst the pins provide amplification of the signal. This is analogous to the interaction between Merkel cells and intermediate ridges within the human fingertip [199], where intermediate ridges help to focus the stress of contact as well as magnifying the signal [39]. The dots move relative to the optical system, due to tactile stimulation, resulting in pixel change. The raw values of all 900 pixels (6-bit resolution) are used as features for tactile perception. This is a completely novel approach compared to past work with the TacTip, where pin deflections were used.

3.5.1 Task: Location and Angle Perception

For high-resolution sensing, using mode-HR, we implement a biomimetic perception algorithm previously shown to achieve superresolved acuity with the TacTip [26]. These methods have successfully been applied to a contour following task [189], and here we re-implement the same methods, applied to a different sensor and a different representation of the data (pixel values instead of pin deflections).

3.5.1.1 Data Collection

For each of the four alternate pin layouts show in Fig. 3.4 we collect two distinct datasets; the first is used for training and the second for testing through offline validation. The robot makes successive taps onto the stimulus edge and records 5 frames with the sensor held statically at the bottom of the tap (~2 mm of compression of sensing surface) for 2.5 s. Over the 5 frames, pixel values for all $N_{\text{dims}} = 900$ are recorded yielding a total of 4500 sensor values. The stimulus used was a 3D-printed circular object (diameter = 107 mm) with a 90° edge, shown in Fig. 3.6.

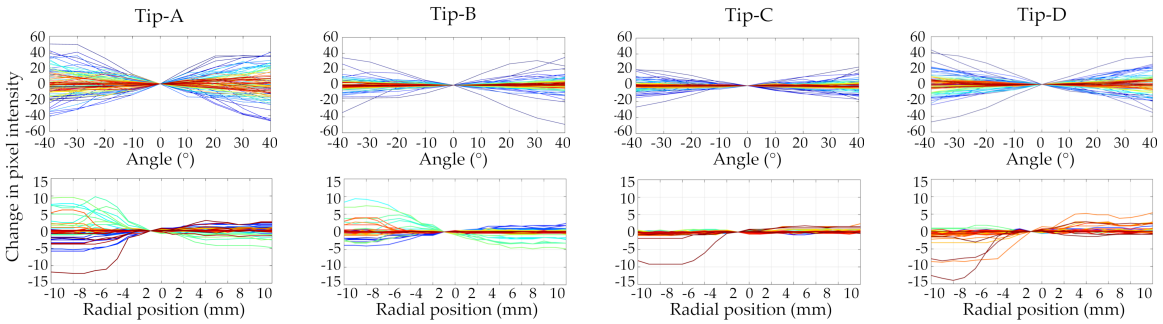

Figure 3.9: Change in pixel intensity for 100 selected pixels, plotted in different colours, against angle (top) and radial distance (bottom) for the four pin layouts shown in Fig. 3.4. The values are plotted as relative to the values at class zero for both angle and radius, ($\theta = 0$, $r = 0$), above and below respectively. The pixels are selected for each plot to display maximum variance.

All dimensions are measured in the sensor frame, the origin of which is located at the sensor's lowest point (see Fig. 3.11). For each dataset, taps are performed at a discrete set of angles and radial positions relative to the stimulus edge (the same for both sets). We use $N_\theta = 9$ angles spanning $-40° \leq \theta_i \leq 40°$ and $N_r = 21$ radial positions spanning $-10\,\text{mm} \leq r_l \leq 10\,\text{mm}$ centred on the edge of the stimulus, yielding a total of $N_{\theta,r} = 189$ positions per dataset.

Data collected with mode-HR for the four tips are shown in Fig. 3.9. Here we show the change in pixel intensity, relative to values at $\theta = 0°$ and $r = 0\,\text{mm}$ above and below respectively, for 100 selected pixels, shown in different colours (pixels selected to display maximum variance). The amount of variance that each pixel value exhibits may be an indication of how useful that pixel is as a predictor. The plots show significant variation in pixel values across the sample range, suggesting that, in general, pixel values may be suitable predictors of θ and r. However, the relationship between (r,θ) and pixel values, s_k for a given pixel, k, is a complex multidimensional function, $s_k = f_k(r,\theta)$, where s_k co-varies with θ and r simultaneously. Thus, we cannot draw a hypothesis from Fig. 3.9 with regards to relative perceptual performance of each tip over the entire range of percepts (r,θ). Likewise, this result is for data collected on a circular edge only: for other shapes the function, f_k, will likely be very different.

3.5.1.2 θ and r Decision Making

For decision making, we adopt a standard 'histogram' likelihood model [248]. The data, d, is considered as a time series of sensor values,

$$d = \{s_k(j) : 1 \leq j \leq N_{\text{samples}}, 1 \leq k \leq N_{\text{dims}}\}, \tag{3.2}$$

j denotes the time sample and k denotes the sensor dimension (pixels). After each test tap, $N_{\theta,r}$ likelihoods are calculated, one for each position, using a measurement model of the training data.

$$\log P(d|r_l,\theta_i) = \sum_{k=1}^{N_{\text{dims}}} \sum_{j=1}^{N_{\text{samples}}} \frac{\log P_k(s_k(j)|r_l,\theta_i)}{N_{\text{samples}} N_{\text{dims}}}, \tag{3.3}$$

constructed by assuming that all pixels and samples are independent, where r_l and θ_i are competing radial and angle position hypothesis respectively. The probabilities, $P_k(s_k(j)|r_l,\theta_i)$ are found with a histogram method applied to training data for each class. The samples, s_k, are binned into equal intervals I_b, $1 \leq I_b \leq N_{\text{bins}}$ over their range (here with $N_{\text{bins}} = 100$). This is transformed to a probability distribution by normalising over the total number of samples in each class membership.

$$P_k(s_k|r_l,\theta_i) = P_k(b|r_l,\theta_i) = \frac{n_{kli}(b)+\epsilon}{\sum_{b=1}^{N_{\text{bins}}} n_{kli}(b)}, \tag{3.4}$$

where $n(b)$ is the total number of sample counts in bin b. Using a constant offset, $\epsilon << 1$, avoids taking the log of zero.

A decision regarding the sensor state is made each time new sensory data becomes available. The decision is made according to a maximum-likelihood criterion.

$$\theta_{\text{dec}} = \arg\max_{\theta_i} P(d|\theta_i) = \arg\max_{\theta_i} \sum_{l=1}^{N_r} P(d|\theta_n, r_l), \tag{3.5}$$

$$r_{\text{dec}} = \arg\max_{r_l} P(d|r_l) = \arg\max_{r_l} \sum_{i=1}^{N_\theta} P(d|\theta_i, r_l). \tag{3.6}$$

Likelihoods, $P(d|\theta_i)$ and $P(d|r_l)$, are found by marginalising the joint likelihood over radial position and angle respectively. We opted to use separate likelihoods instead the joint likelihood because on θ predictions, for example, it is likely to be more accurate due to averaging across the r dimension.

3.5.1.3 Results

To test the perception, we perform offline validation, computing average angle and radial errors, e_θ and e_r, with 10,000 samples randomly selected from the test set.

We computed error values as absolute discrepancies in perceived angle and radius with the ground truth, averaged over multiple test runs, as a function of angle and radial position; $e_\theta(\theta,r) = \langle |\, \theta_{\text{dec}} - \theta \,| \rangle$ and as a function of radial position; $e_\theta(r) = \sum_{i=1}^{N_\theta} e_\theta(r,\theta_i)/N_\theta$.

All four tips achieved good perceptual accuracy. These error values are shown for tips A and B in Fig. 3.10.

Table 3.2: Mean angle and radial errors in an offline perception task. Values were computed ignoring the radial position classes $-10 \leq r \leq -8$.

	Mean angle error, e_θ (°)	Mean radial error, e_r (mm)
Tip-A	0	0.06
Tip-B	1.4	0.07
Tip-C	0.4	0.01
Tip-D	1.5	0.06

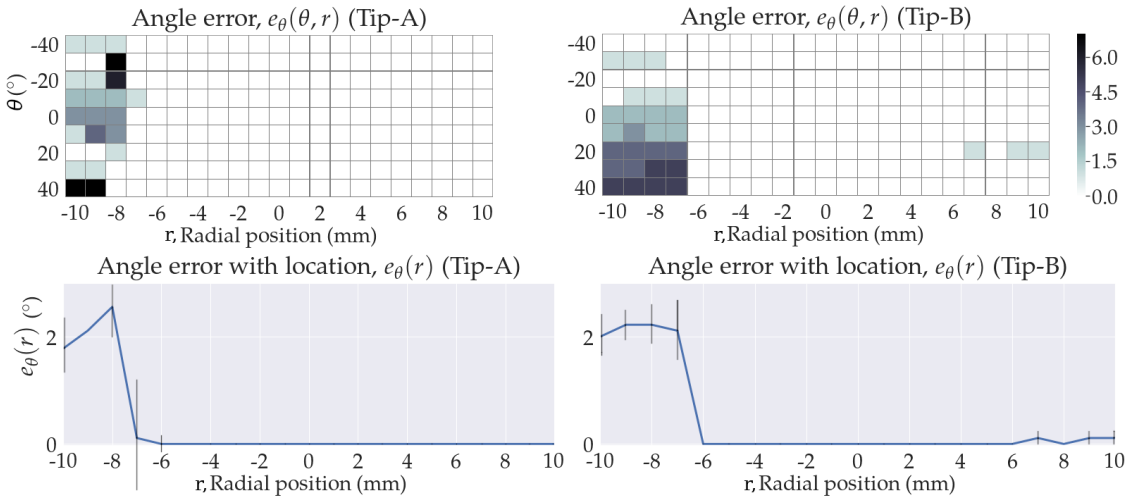

Figure 3.10: Angular perception errors, with angle and radial position, $e_\theta(\theta, r)$ (top) and with radial position, $e_\theta(r)$ (bottom) for tips A and B. Above errors are a heat map: 70° (black) - 0° (white).

We note that there is near perfect angle perception at
positions $r \geq -8$ mm. Below -8 mm the sensor taps are in free space, hence perceptual performance here is determined by chance. Error bars show the standard error in the mean.

Tip-B experiences some loss of perceptual accuracy as contact is further onto the stimulus. We interpret this as a consequence of the edge signal becoming reduced with respect to normal deformation as a result of the tip pressing on the flat top of the stimulus. Average angular and radial errors, for all four tips, were below 2° and 0.1 mm respectively. These values are shown in Table 3.2. We omitted results at positions $r < -8$ mm since these taps were in free space.

3.5.2 Task: Contour Following

In the presented exploration task, the robot is required to follow the edge of an unknown object using only tactile feedback. Tactile contour following requires both sensitive and robust perception. It is therefore an ideal task to measure the performance of the high resolution sensing modality.

3.5.2.1 Data Collection and Training

For training the robot, we use the same training dataset as in Section 3.5.1.1. The procedure for building the likelihood model is described in Section 3.5.1.2.

3.5.2.2 Control

We implement a 'tactile servoing' approach for robot control throughout the contour following task which was previously developed for implementation with the TacTip [189]. Here we have adapted the methods for application with this specific sensor.

The robot is controlled through a perception-action cycle. The sensor is tapped onto the stimulus edge where 5 frames are recorded. The stimulus edge angle and radial position, are then perceived, in the sensor frame, according to the procedure described in Section 3.5.1.2. This perception informs the actions to select via a deterministic control policy.

The method for controlling the robot involves an action selection policy, where the robot attempts to maintain an optimal position for perception. For the purpose of tactile contour following, action selection defines two procedures: *(i)* tactile servoing over θ, which aims to maintain a constant angle, θ_{fix}, relative to the stimulus edge and *(ii)* tactile servoing over r, where the sensor relocates perpendicularly to the perceived edge towards a pre-set fixation radial displacement, r_{fix}. The two policies are described by the following two equations respectively:

$$\Delta\theta = \pi_\theta[P(d|r_l,\theta_n)] = \lfloor g_\theta(\theta_{\text{fix}} - \theta_{\text{dec}})\rfloor_i, \tag{3.7}$$

$$\Delta r = \pi_r[P(d|r_l,r_l)] = \lfloor g_r(r_{\text{fix}} - r_{\text{dec}})\rfloor_l, \tag{3.8}$$

where g_θ and g_r are the angular and radial gain factors respectively (set to 0.5 or 1 in previous work [189] as these were found to achieve good results with the same shape and exploration step size through experimentation.) and $[.]_l$ and $[.]_i$ shows that the action is rounded down to the nearest class. Here the fixation points are chosen as the centre of the perceptual ranges, $\theta_{\text{fix}} = 0$ and $r_{\text{fix}} = 0$. The perceived angle and radial position, θ_{dec} and r_{dec} respectively, are obtained from the decision making process described in Section 3.5.1.2. Hence, by performing a sequence of perceptions followed by informed actions we implement a perception-action cycle. Fig. 3.11 shows how dimensions r, θ and q relate to the sensor frame and stimulus.

Supplementary to action selection is an exploration policy. Also defined by perception, the robot moves the sensor in a direction tangential to the perceived edge angle by a fixed amount, Δq, here we use 3 mm as with [189]. It is via this exploratory move, that the robot traces the edge of the stimulus object.

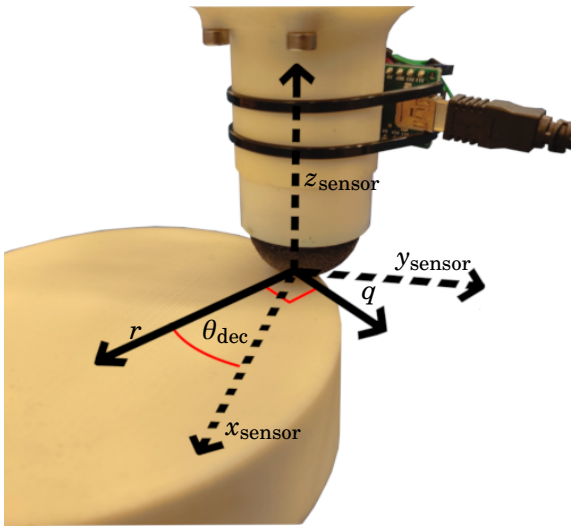

Figure 3.11: Diagram of dimensions in control algorithm. Sensor frame co-ordinates axis are shown as dashed lines; θ_{dec} is the perceived edge angle in the sensor frame; r, perpendicular to the perceived edge, is servoed over to maintain a centralised position of the edge in the sensor frame; q, perpendicular to r, is the exploratory direction.

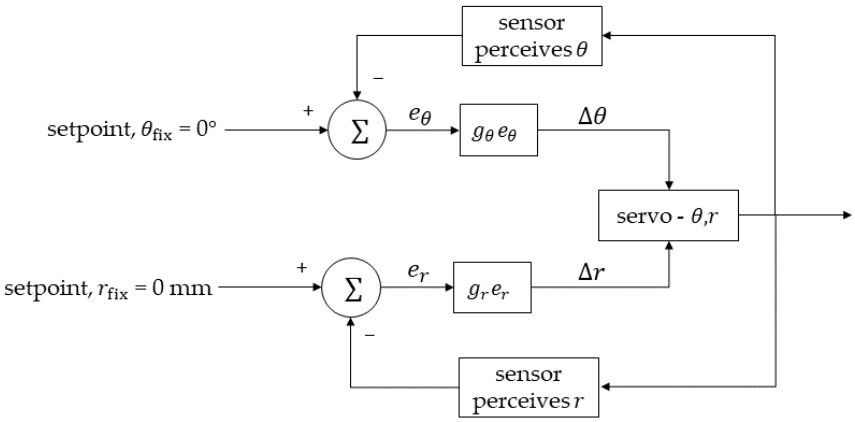

Figure 3.12: Integrated proportional controllers for θ and r. e_θ and e_r are the angular and radial position errors respectively.

The procedure described in Equations 3.7 and 3.8 can also be interpreted as two parallel proportional controllers. A diagram of this interpretation is shown in Fig. 3.12.

In this interpretation, the control loop is implemented once per tap and followed with the exploratory move, Δq. The perception blocks are implementations of the decision making procedure described in Section 3.5.1.2. Angle and radial error values, e_θ and e_r, are computed as the discrepancy of perceived values from their respective pre-set fixation points. The servo block transforms control signals, $\Delta\theta$ and Δr, from the sensor frame into the robot base frame and performs relevant point-to-point Cartesian moves via the robotic system described in Section 3.5.1.2.

3.5.2.3 Results

To test the ability of our sensing methods, the robot is required to follow the contour edge of a circular object (diameter = 110 mm) using the same training dataset as used in Section 3.5.1.3.

A radial fixation point centred on the range, $r_{fix} = 0$ was chosen. This location is validated

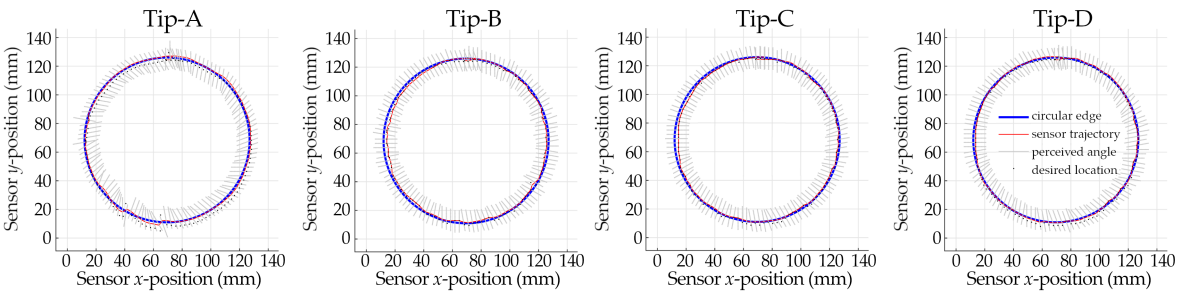

Figure 3.13: Trajectories for a contour following task around a circle (diameter 110 mm) (blue curve) for four pin layouts. Radial and angular gains are 0.5.

by the results from the perception task (Section 3.5.1.3), since perceptual performance was good here. We performed the task with all four tips and the results are shown in Fig. 3.13.

Good performance was found with radial and angular gains $g_r, g_\theta = 0.5$ and exploration steps $q = 3$ mm. All four tips successfully completed the task and the circular edge is followed to a high degree of accuracy. Visual inspection of the trajectories shows minimal variation in performance between the tips. In terms of sensor trajectory (how closely the sensor followed the circular edge), Tips B, C and D were all comparable and very good, whilst Tip-A performed noticeably worse than the competition.

Table 3.3: Mean positional error for the four alternate pin layouts in the contour following task.

	Mean positional error, e_d (mm)
Tip-A	2.9
Tip-B	2.0
Tip-C	2.8
Tip-D	2.4

The mean positional error for each tip in the contour following task is shown in Table 3.3. These were computed as the mean of the Euclidean distance of the sensor from the desired circular trajectory at each tapping position. The perceptual accuracies appear to confirm the performance in contour following. Interestingly, this result counters the angular perception results from Section 3.5.1.3, suggesting that, whilst Tip-B may under-perform when perceiving angles that have previously been observed, this tip was able to generalise to unseen stimulus resulting in better contour following.

3.6 Combined Modality Control

To demonstrate the benefits gained from dual-modality we design a task where the robot uses both modes. Tip-B was chosen for this task: as in Section 3.4.1.3, it was found to be the most suitable for use with mode-HS, and shows comparable performance with other tips for perception in mode-HR (Section 3.5.1.3) as well as contour following tasks (Section 3.5.2.3).

The robot must follow the edge of a previously unseen stimulus, trained only with the circle shape used for previous experiments. A non-uniform volute shape (see Fig. 3.6) is used, where the radius of curvature varies from 20 mm to 50 mm.

Low threshold contact contact detection is believed to signalled by one or very few FA

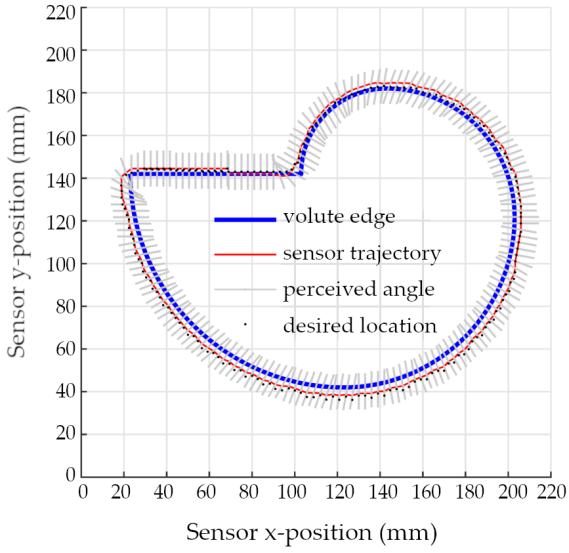

Figure 3.14: Trajectory for contour following task on a non-uniform volute (blue curve). Radial and angular gains are set to 0.5.

afferents and/or PC afferents [51]. Mode-HS
shares characteristics with the possible signals thought to encode contact detection: high speed and spatially coarse. We use mode-HS, therefore, to detect contact on the volute shape whose exact position is unknown to the robot. First, the robot moves the sensor down to a rough approximation of the centre of the test object. The surface quality measure is monitored and initial contact is detected by applying a threshold criterion.

To successfully apply training data to contour following with this unseen object, the sensor must attain the same relative height. This is achieved with the use of mode-HS. After contact is detected, the sensor height is modulated to the correct depth according to Equation 3.1; i.e., by applying contact depth control (Section 3.4.1). The robot then locates the edge of the test object by monitoring surface quality whilst moving in the x, y-plane. The edge is detected when surface quality rises above a threshold.

Once the edge is detected, the robot switches to mode-HR and follows the contour according to the procedure described in Section 3.5.2.2. Results of contour following on the volute, with an automated dual-mode approach to depth and edge location, are shown in Fig. 3.14. A video of the robot performing combined modality control is available on YouTube [255].

The robot successfully detects initial contact and locates the edge of the volute using mode-HS. The robot then correctly follows the contour, suggesting that the depth is modulated accurately and general training data can be applied to the methods in order to follow the contour of an unknown, non-uniform shape using the second modality.

3.7 Concluding Remarks

In this chapter, we presented a novel, dual-modal, optical tactile sensor, without image processing or a high-resolution imaging system. The sensor was tested in a series of tasks, aimed at assessing its ability to use a high-speed modality for contact detection and depth modulation and a second, high-resolution modality for perceiving angle and radial position relative to an edge. We examined the robustness of the high-resolution perception in a tactile contour following task, using the methods developed in [189]. Finally, we presented a task requiring both modalities, where the robot follows the contour of a previously unseen object, where the exact object location and depth was unknown by first detecting contact, modulating to the correct height and locating the edge using the high-speed modality before switching to the high-resolution mode for contour following.

The sensor achieved good accuracy in both the depth control and angle and radial position perception tasks. The high-resolution sensing was robust enough for successful contour following around the circle and this generalised to the non-uniform shape after using the high-speed modality to find the object's height within the robot's base frame and locate the edge.

The presented high-resolution mode was leveraged for contour following with comparable accuracy to the TacTip using the same methods for control and perception [189]. This is a

very interesting result since the high-resolution mode used here features a significantly lower resolution camera system (30x30) than the TacTip (640x480). This has significant implications for camera-based optical tactile sensing, potentially reducing manufacturing costs and aiding miniaturisation.

We compared the performance of four alternate pin layouts in the depth control, perception and contour following tasks. Interestingly, there was little variation in performance between the four pin layouts within the high-resolution mode when perceiving angle and radial position. This was a surprising result; however, our study was limited to concentric patterns of pins which, whilst necessary for the TacTip, could be replaced by an arbitrary pattern in the case where raw pixel values are used as features. Hence, further study may elicit a more optimal pin design. The pin layout had a large influence on the high-speed sensing modality: Tip-B showed the most sensitivity in the image quality statistic to tactile stimulation. Image quality is a measure of the number of features present in the frame [246]. It appears to be maximised when pins are at the optimum distance from the lens and reduces as the pins move out of focus. We suggest that Tip-B has the best balance between pin spacing, number of pins and pin size for the image quality algorithm to detect each pin as a separate feature. The ADNS-3080 may struggle to separate the pins of Tip-A, for example, as separate features, whilst tips C and D have fewer pins, yielding a lower overall value in image quality.

Here we focus on using the high-speed mode for contact detection, similar to to a hypothesised function of FA and/or PC afferents [51] and a depth control sensor which is necessary to achieve robust tactile sensing in unknown environments. It is known that in nature reflex action serves to minimise the time taken to respond to dangerous situations [31]. We argue that depth control via the high-speed modality provides a valid demonstration of the ability to perform artificial reflex action since it demonstrates the ability to sense and respond to adverse conditions of high pressure due to normal forces.

Previously we made the distinction between **multi-modality** (multiple modes can be used at once) and **dual-modality** (each mode must be used in isolation). The presented sensor is dual-modal which presents a significant drawback compared with other multi-modal tactile sensors; e.g., the BioTac [7, 182, 256] and other less prominent multi-modal fingertips [168, 186, 257] and tactile skins [184, 185]. Whilst these sensors have the benefit of *multi* over *dual*-modality, there are few examples in which both modalities have been leveraged concurrently, and fewer still where tactile sensing is required for real-time applications. In particular, the BioTac has been extensively employed in 'uni-modal' tasks [183, 194, 210, 258, 259] where only one of its three modes is used. There are some cases where the multiple modalities of the BioTac [7, 260] and other multimodal fingertips [186] have been leveraged for object recognition tasks, however, object recognition does not necessarily require simultaneous sensing from multiple-modalities: each modality can be used sequentially and subsequently fused; i.e., in a dual-modal fashion. In-fact, even in real-time control tasks, where *multi-modal* sensing may be useful, the BioTac

has been used in a *dual-modal* fashion to estimate force for grasp control and sequentially detect slip events [244].

True multi-modality has been leveraged in real-time control tasks in relatively few cases. For example, a multi-modal tactile skin, is used in a human-robot interaction experiment where pressure and proximity measurements are fused to allow a robot arm to react to external contact [261]. Grip force and slip events are simultaneously monitored by using pressure arrays and accelerometers within the PR2 parallel jaw gripper in a human-inspired grasp control framework [262].

It is difficult to directly compare the aforementioned studies with the presented work due to the differences in the testing scenarios and the intended function of the different modalities. It is intuitively obvious that *multi-modality* has benefits over *dual-modality*, however, we have seen that in only very few cases true multi-modality is leveraged. Furthermore, we believe it is highly-likely that the presented sensor surpasses aforementioned multi-modal sensors in terms of spatial resolution, owing to its camera-based optical tactile sensing. Also, the presented sensor is an extremely low cost option for dual-modal tactile sensing compared with the presented multi-modal alternatives.

It is worth considering the work of Knoop and Rossiter (2013) [187], which uses similar technology to that developed in the present study to perform an approximate and fast force estimation in 'reflex-mode' (analogous to mode-HS) and a more accurate but slower force estimation in 'explore-mode' (analogous to mode-HR). Essentially, the work in [187] demonstrated that an extremely low cost and low resolution image acquisition chip could be used for dual-modal force sensing. The work presented here can be seen as a significant development where we demonstrate that the same style of image acquisition chip can also be used for approximate and fast robot control in high-speed/reflex mode and, additionally, accurate and robust control is achieved using the high-resolution/explore mode which is comparable with more expensive and higher-resolution tactile sensors; e.g., the TacTip [189].

CHAPTER 4

ARTIFICIAL TACTILE AFFERENTS FROM MARKERS & VIBRATIONS

In this chapter we present three hardware modifications to the TacTip: *(1)* a novel marker arrangement, designed to more closely mimic the density of SA-I and FA afferents in the human fingertip; *(2)* an artificial fingerprint for the purpose of vibration amplification; *(3)* an additional modality, a *vibrational channel*, which we suggest as an analogue of the natural PC channel.

We propose a novel method for extracting features from the TacTip camera image, designed to model rate response of natural SA-I and FA afferents.

Finally, we perform a set of experiments designed to understand the likeness of our artificial channels to their natural analogues and assess their capacity for mediating tactile information pertaining to texture.

4.1 Background

As reviewed in Section 2.1.1.1, humans possess four discrete tactile afferent types for discriminative touch. The properties of these primary afferents, in terms of morphology and response characteristics, have evolved for the perception of a diverse range of tactile dimensions (Section 2.1.3). It is thought that each tactile dimension is mediated in a specific way, employing different combinations of afferents and varied encoding within their populations. We suggest that, for robots to achieve the competency seen in human touch, engineers must consider all tactile dimensions. One approach to developing the hardware capacity for this is to develop sensors that model natural tactile afferents.

Many of the available tactile sensors are inspired by the tactile channels of human touch. For example, TacTip markers have been proposed as analogous to Merkel cell complexes (SA-I) [12]; Romano et al. (2011) [262] loosely mimicked SA-I, FA and PC channels from capacitive array

outputs of 5x3 taxels and accelerometers; and the BioTac [256] has been suggested to possess artificial channels analogous to SA-I and PC. In some cases, natural tactile channels are directly modelled to produce spike based outputs; e.g., using biological models of SA-I [263, 264] and FA [264, 265] neurons, in a subset of tactile sensors described as 'neuromorphic'.

Unlike the aforementioned studies with bio-inspired tactile sensors [12, 256, 262], where the technology is inspired by biology, here we attempt to directly model firing rate of SA-I and FA natural tactile channels deriving this 'artificial code' from TacTip marker positions. Additionally, we directly compare outputs of our artificial SA-I and FA afferents with natural analogues under the same stimulation conditions.

Our approach is also different to neuromorphic tactile sensors in that we do not attempt to model the precise output of biological neurons; i.e., spike trains. As mentioned, the continuous output of our artificial afferents are proposed to model firing rate in their natural counterparts; i.e., we assume a rate code for single neuron firing of SA-I and FA afferents. By assuming this abstract model of neural code, our approach is simpler than neuromorphic methods and is more easily applied to conventional machine learning models for decoding spatio-temporal signals.

We make a number of modifications to the TacTip. Most notably, the TacTip is augmented with an additional modality, *vibration channel*, via a microphone. We propose this may form an analogue of the natural PC channel since it is sensitive to the types of tactile cues to which the PC channel is most responsive; e.g., high frequency vibrations [44].

We are primarily interested in touch sensing mediated through physical interaction with the environment. Therefore, we do not consider SA-II afferents (Ruffini endings), as these are generally believed to be primarily responsible for proprioceptive sensing via mediating information about skin stretch [44, 46]. Indeed, SA-II afferents are rarely considered in aforementioned studies into human sensing of key tactile dimensions (Section 2.1.3) although some evidence suggests that those located around the nail-bed are responsible for sensing shear force direction [266].

We perform three experiments with the aim of ascertaining the degree of likeness of artificial tactile channels to their natural analogues. Firstly, we qualitatively assess the response of three engineered channels to simple sustained and transient normal pressure at different speeds. We then compare output of artificial SA-I and FA afferents to neurophysiological readings from monkey SA-I and FA afferents when stimulated with ridges of varying widths, and edges of varying separation. Finally, we examine the response of all three artificial channels to scanning with raised bump textures of varying bump diameter and separation with the aim of understanding potential artificial codes for texture discrimination.

4.2 Sensor Design

Unlike in the preceding chapter, where the dual-modal sensor was heavily inspired by the TacTip, here the presented sensor is an extension of it. Thus, in this chapter the sensor is referred to as a TacTip. We also reiterate the difference between **dual-modality** and **multi-modality**. Where, in the previous chapter, the presented sensor was dual-modal, here the work is towards enabling multi-modality for the TacTip such that it can simultaneously record from multiple channels.

4.2.1 The TacTip

Here we describe the design and manufacture of the latest iteration of the TacTip, an established sensor on which the novel contribution in this chapter is built upon. For a conceptual overview, including biomimicry, please refer to Section 2.2.1.1 of the literature review.

The TacTip (see Fig. 4.1) is fully 3D-printed and comprised of two primary sections: a rigid body, printed in ABS by a Fortus PolyJet printer from Stratasys and a soft tip printed by a multi-material FDM printer, Objet, also from Stratasys. Modularity enables us to easily experiment with different tip designs, additionally, the tips can be easily substituted if damaged rather than replacing the whole sensor. The two elements are fastened via a bayonet mounting system.

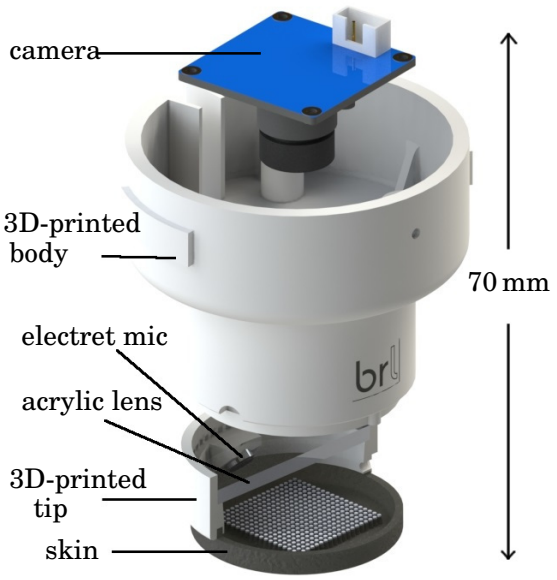

Figure 4.1: Exploded view of the modified TacTip design. The dimension (70 mm) refers to the 'unexploded' size.

The tip is comprised of two materials printed as a single part: a rubber-like 'skin' (Tango Black+ (Shore A 26-28), Stratasys) and rigid rim (Vero White, Stratasys). On the inside surface of the skin, a set of flexible pins are printed in Tango Black+. On the tip of each pin a Vero White marker is printed. The generic TacTip has 127 markers forming a regular hexagonal pattern (see Fig. 2.12). A clear acrylic lens is glued to the inside of the rim which creates a cavity between the inside of the skin and the lens. The cavity is injection filled with a clear silicone gel (RTV27905, Techsil UK (~Shore OO 10)). The gel provides a stiffness to the tip which helps to minimise hysteresis whilst still enabling compliance.

The TacTip body's primary function is to house electronic components and to serve as a mount for auxiliary robotic systems; e.g., robot arms. To capture images of the white markers, we use

a 1080p USB camera (ELP cctv), capable of capturing frames at up to 100 Hz. It is fastened to the body via 4 M3 screws. The length of the shaft (~20 mm) is chosen to provide full view of the markers through the wide-angle lens. At the base of the shaft is a PCB ring containing a set of 6 white LEDs which illuminate the markers for detection by the camera.

4.2.2 Modifications

4.2.2.1 A Tactile Vibrational Channel

In accordance with the aforementioned motivations, we facilitate multi-modality for the TacTip by leveraging acoustic vibrations that occur within the gel during dynamic stimulation. To detect these vibrations, a small electret microphone [267] (diameter = 6 mm) is fitted to the inside of TacTip's Vero White rim (see Fig. 4.2). During assembly, the microphone is press fitted into a cavity prior to fixing the acrylic lens. The microphone's cabling is routed through a hole in the side of the tip. After glueing the lens, the tip is filled with silicone gel as normal (Section 4.2.1). The Gel is in contact with the mic.

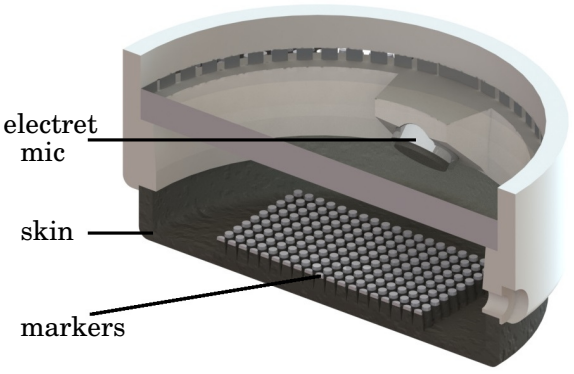

Figure 4.2: Cut-away view of the modified TacTip.

In principle, a physical interaction between the TacTip's 'skin' and the stimulus will induce vibrations which propagate through the silicone gel. Which are converted to a voltage by the microphone.

An electret microphone is a type of electrostatic capacitor-based microphone. Within the microphone, a pair of parallel charged plates have an inherent capacitance which varies according to the distance between the plates. One plate is able to move as a consequence of local pressure changes thus producing a variable potential difference across the capacitor. Output from these devices is typically on the order of millivolts.

For interfacing with a PC we amplify this signal to line-level (V_{pp} = ~2 V) using an inverting op-amp circuit. The gain was adjusted such that minimal clipping occurred during intended styles of stimulation to the TacTip. The pre-amplifier circuit is housed in a grounded die-cast enclosure (see Fig. 4.4) to minimise the effect of electrical noise. The amplified signal is connected to a PC soundcard via a mono co-axial audio cable.

4.2.2.2 Tip Shape

Rather than the traditional hemispherical shape, here we have opted for a flat skin. The rationale behind this modification was to reduce the physical distance between the point of interface with stimuli and the microphone. We believe that acoustic vibrations are likely to attenuate as they travel through the silicone. Therefore, by moving the skin closer to the microphone, we hope to maximise the level of signal that is detected. In practice, after the tip has been filled with gel the skin is no longer completely flat, but rather forms a slight 'bulge'.

4.2.2.3 Marker Layout

The TacTip design here uses 361 markers rather than the usual 127 which is likely to improve spatial resolution. Additionally, the pins are arranged in a simple 19x19 square array with a separation of ~1.5 mm (see Fig. 4.2), rather than the traditional hexagonal pattern. We use a regular square array so that information encoded in the relationship between movements of neighbouring markers (spatial coding) can be more easily identified.

4.2.2.4 Artificial Fingerprint

As described in Section 2.1.1, there are a number of theories as to the purpose of epidermal ridges and fingerprints. In particular, Scheibert et al. propose that the human fingerprint may amplify the vibrations produced at the interface of the skin and stimulus which could be beneficial for texture perception [59]. It is therefore interesting to investigate the affects of an artificial fingerprint when tasking our robot to perceive texture.

Researchers have previously observed enhanced edge localisation performance on small scale features such as sharp edges with the TacTip when endowed with an artificial fingerprint [199]. Here we leverage the design concepts realised by Cramphorn et al. (2017) [199] and produce a similar modification to our multi-modal design. Fig. 4.3 shows a comparison of the standard 'smooth-tip' with the modified 'fingerprint-tip' version. Two main design modifications are implemented to mimic *i)* a fingerprint and *ii)* epidermal ridges. *i*) The outside of the skin is augmented with small nodules (diameter ~2 mm) which form the physical analogue of papillary ridges (fingerprint). The nodules are made from the same material as the skin and thus are also compliant. The pattern nodules on the skin exterior mirrors that of the pins on the interior: nodules are located directly below every second pin. *ii)* Each pin contains a rigid Vero White 'core' that is mechanically fused with its respective white marker. The cores extend through the pin to approx ~0.5 mm from the outside surface of the skin. The purpose of the cores is to enhance the stiffness contrast between the pins and the silicone gel. This stiffness contrast is designed to mimic the proposed function of the stiffer epidermal ridges [39] (analogue of TacTip pins) which reach down into the softer dermis layer terminating in a Merkel cell complex (analogue of TacTip markers) (see Fig. 2.1). Both nodules and cores are printed in the same single part as the rest of the tip.

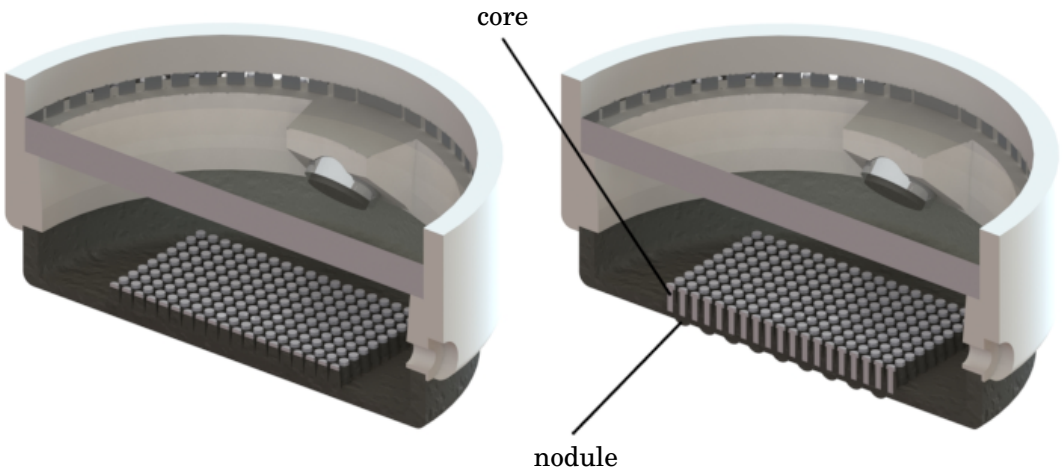

Figure 4.3: *Left*: Modified smooth tip. *Right:* fingerprint tip.

4.3 Robotic System and Software Structure

A high-level block diagram of the overall software structure is shown in Fig. 4.5.

We interface with the TacTip via a Python Sensor class which contains methods for recording video data from the USB camera and vibration data from microphone. For recording video data, we utilise the v4l2capture library [268] which is an open-source Python extension for capturing video using the video4linux2 driver (using this system we attain a maximum frame rate of ~90 fps). We leverage the open-source sounddevice Python module [269] for recording audio data, which provides bindings for the PortAudio open-source, audio I/O library written in C.

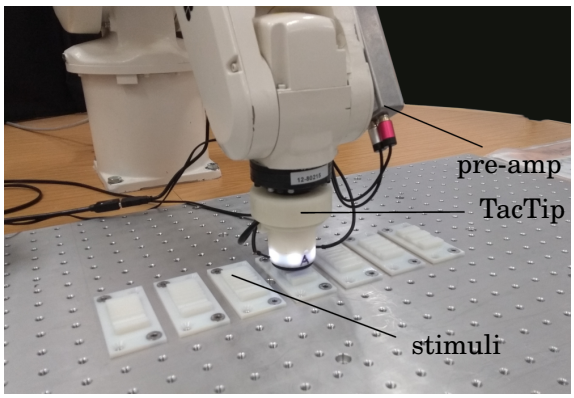

Figure 4.4: TacTip mounted as an end-effector to the ABB robot arm, collecting data on a set of grating stimuli.

We enable the sensor to interact with its environment with the use of a six degree-of-freedom robot arm (IRB120, ABB Robotics) to which our sensor is mounted as an end-effector (see Fig. 4.4). The arm can precisely and repeatedly position the sensor (absolute repeatability 0.01 mm). A custom 3D-printed mount is bolted to the rotating (wrist) section of the arm to which the sensor is attached via a bayonet mechanism.

For controlling the arm we leverage the open_abb open source software project [254]. open_abb

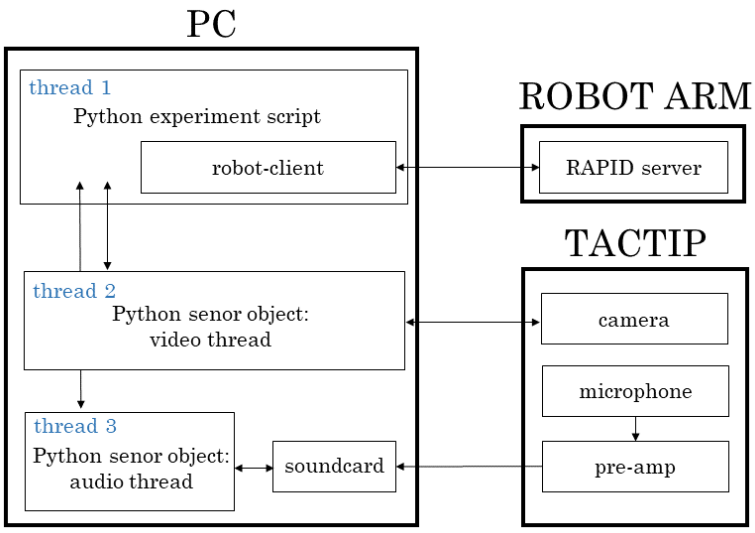

Figure 4.5: Diagram of the robotic system: high-level software structure and abstract hardware interfaces.

provides a Python client API in the form of a Robot class which provides methods to perform operations such as Cartesian moves, joint moves, setting the speed etc. open_abb also provides a server script written in RAPID, the ABB controller native language. This runs on the ABB control unit in a continuous loop. When these methods are implemented, commands are sent, over a network, to the RAPID server via TCP and placed on a software buffer. The RAPID script will process the data within this buffer and command the robot accordingly.

For collecting data, we produce an experiment script. Written in Python, it instantiates a robot-client object of the Robot class and a sensor object of the Sensor class. Importantly, each time we require either video or audio data, a separate thread is created for each channel, this enables us to asynchronously record tactile data whilst simultaneously moving the robot by calling methods on the robot-client object.

4.4 Artificial Afferents

It has been proposed that marker positions are analogous to SA-I mechanoreceptors (Merkel cells) [12] and, more recently, that marker velocities could represent FA mechanoreceptors (Meissner corpuscles) [270]. Traditionally, x and y marker positions have been employed as tactile features with the TacTip and recently raw images have been used to train convolutional neural networks for automatic feature detection [6]. In contrast to these techniques we develop a more bioinspired method for feature extraction that aims to model SAI and FA channels, with the intention that these features closely mimic readings from equivalent natural tactile afferents. Additionally, we have engineered a novel tactile vibrational channel which is inspired by the Pacinian system in

human touch.

In the human tactile system, groups of, or single, mechanoreceptors of an individual type are innervated by single afferent nerve fibres. Thus, each afferent corresponds to a particular tactile channel (SA, FA or PC) and each afferent has an effective receptive field owing to the arrangement of its associated mechanoreceptors and the characteristics of the mechanoreceptors themselves (Section 2.2).

The firing rate of these afferents is sometimes considered a fundamental neural code for single afferents (rate code) and may be the building block upon which more complicated coding regimes; e.g., spatial, are abstracted (Section 2.1). It is this concept which motivates implementation of a novel feature extraction technique: a single value per tactile channel should be attained that is analogous to afferent firing rate, such that abstract biological coding regimes (e.g., spatial and temporal) may be more closely mimicked. It is possible that in nature precise inter-spike timing is important for encoding many tactile stimuli including texture [271]. We are unable to model this type of coding regime due to the form of the data provide by the camera system, specifically its discrete sample rate. This limitation is discussed in more detail in the Concluding Remarks, Section 4.8.

4.4.1 SA-I Afferents

SA afferents refer to a class of afferents with slow adaptation rates (see Section 2.1.1.1). That is, they respond to sustained pressure or deformation of the skin as their associated mechanoreceptors (Merkel cells) are tonic. Type-I implies that the afferents densely innervate skin regions and exhibit small receptive fields relatively with well defined borders [30]. They are therefore associated with high spatial acuity.

In our approach, SA-I firing for each channel is modelled as the Euclidean distance of a marker from its at-rest position. Consider a sustained stimulus, the deformation of the tip will remain consistent and therefore so will the positions of the markers. Thus, response, computed as Euclidean distance, is slowly-, or in practice, non-adapting. Just as biological afferents have receptive fields that are spatially arranged on the skin, the artificial afferents are spatial arranged in a 2D array according to their corresponding marker's rest position. There are $n_{\text{markers}} = 361$ channels, where n_{markers} is the number of markers within the TacTip. This yields a density of ~40 afferents per cm^2. We consider the arrangement of markers to be indicative of high innervation density, thus, modelling type-I mechanoreceptors. In biology, the exact locations and shapes of receptive fields of individual afferents in a population are used in combination with their firing rates to spatially encode stimulus percepts [43]; i.e., the spatial modulation of afferent firing provides a spatial code for the stimulus. In drawing this analogy, we cement the idea that in robot touch, stimulus properties may also be encoded spatially, via the pattern of Euclidean distances over the tip.

4.4.2 FA Afferents

As with SA-I, FA afferents are associated with type-I mechanoreceptors, accordingly, they densely innervate the skin [42]. Therefore, in terms of arrangement of receptive fields, can be modelled by some physical quantity associated with TacTip markers.

In contrast to SA-I, FA afferents are rapidly adapting, meaning their firing tends to rapidly decrease when subject to sustained stimulus. In fact, it is believed that these afferents are particularly sensitive to the velocity of the skin within the receptive field [30] and thus tend to fire whenever skin is moving. We model this behaviour as marker speed, which is inherently transient; i.e., will be zero when the stimulus is sustained and positive when the stimulus changes. This model is similar to most transduction models of FA firing, where the first derivative of pressure is used as the primary input to some biological neuron model [265, 272, 273].

4.4.3 Vibrational Channel

PC afferents are rapidly adapting of type II. Their innervation densities are considerably lower than that of SA-I and FA afferents [44] and each afferent is terminated by a single Pacinian corpuscle which are sparsely located deep in the subcutis (see Fig. 2.1). Consequently, they have large receptive fields with obscure boundaries. PC afferents are sensitive primarily to high frequency vibrations (50-500 Hz) and show extremely low amplitude thresholds (1 μm) at frequencies in the range of 100-300 Hz [56].

We have augmented the TacTip with an additional acoustic detection modality, *vibrational channel*, by means of an electret microphone (see Section 4.2.2) which mimics a single Pacinian corpuscle or PC afferent. We propose that this may mediate the type of tactile cues of which PC afferents are particularly sensitive. In design, our system is analogous to the form of the natural PC channel in two ways: Firstly, the microphone innervates the TacTip sparsely relative to markers (artifical SA-I and FA afferents). Secondly, it is located deep within the silicone ('flesh') of the TacTip. Additionally, the information acquired by microphones is, in general, crudely similar to the types of dynamics that the natural PC channel is thought to be preferentially sensitive in three ways: Firstly, they are responsive to physical vibrations (acoustics), which are transient by definition (rapidly adapting). Secondly, the electret microphone considered, has an operating frequency range (10-40 kHz) which includes the frequencies at which the Pacinian system is maximally sensitive. Lastly, the electret microphone considered exhibits extreme sensitivty to low amplitude vibrations.

Importantly, in our approach, whilst we suggest SA-I and FA afferent firing rates may be modelled by marker displacement and speed respectively, we do not do the same for the PC and vibration channels. The vibration data collected via the microphone is a representation of the raw physical vibrations occurring within the TacTip's silicone 'flesh', therefore it is a pure representation of the stimulus. Although findings suggest that the frequency composition of spike trains in PC afferent response matches skin vibration [129], since the vibrational channel

is a completely novel sensory modality for the TacTip, to extract features of the raw physical vibrations that represent some encoding of this spike train is beyond the scope of this work. Instead we examine vibrational data, from which we hypothesise possible coding of the physical stimulus; i.e., which vibrational cues; e.g., frequency, amplitude etc., are linked to relevant percepts and how these may be encoded in a potential artificial afferent. For a definition of the term *cue*, the reader is referred to (Section 2.1.2 of the literature review).

4.5 Experiment: Data Visualisation

An initial experiment is performed for the purpose of visually inspecting output from our artificial afferents at different stages of a basic press and lift contact. Here we hope to gain an understanding of the response characteristics of all three tactile channels, qualitatively compare each channel and make judgements of their similarity to natural afferents.

4.5.1 Data Collection

The robot begins with the TacTip in free-space at ~2 mm above a flat 3D-printed surface. Recording from the TacTip webcam and microphone is initiated and the robot simultaneously moves the TacTip down until at a compression of ~2.5 mm. At the bottom of the tap the robot waits for 3 seconds before returning to its initial position and terminating recording.

We repeat the data collection procedure four times: once for each tip variant (smooth and fingerprint) and with the robot moving at two speeds, 3 and 10 mms^{-1}.

4.5.2 Feature Extraction

In the present chapter, all feature extraction, and subsequent processing, is performed off-line.

SA-I and FA afferents are derived from marker dynamics. Marker positions are tracked, frame-by-frame, with a simple blob detection algorithm implemented using OpenCV in Python [274]. SA-I firing at frame i, for marker n, $\mathrm{SA}_{n,i}$ is computed as the Euclidean distance between marker position at frame i and an initial at-rest frame, $i = 0$.

$$\mathrm{SA}_{n,i} = \sqrt{(x_{n,i} - x_{n,0})^2 + (y_{n,i} - y_{n,0})^2}, \tag{4.1}$$

FA firing at frame i, for marker n, $\mathrm{FA}_{n,i}$ is computed as the Euclidean distance between marker position at frame i and the preceding frame, $i-1$, which is equivalent to the difference between SA-I firing between adjacent frames.

$$\mathrm{FA}_{n,i} = \mathrm{SA}_{n,i} - \mathrm{SA}_{n,0}, \tag{4.2}$$

There are N_{markers} =361 markers in total, so the computations in equations 4.3 and 4.2 are made 361 times per frame: once per artificial afferent.

Traditionally, marker x and y positions have been used as features [26, 189]; i.e., two channels per marker. Artificial SA-I afferents presented are essentially a dimensionally reduced representation of the same data: only the scalar quantity of each marker's displacement is retained. Whilst a vector displacement contains more information and enables measurements of stimulation that Euclidean distance cannot; e.g., circular movements, the aim of the present study is to produce a single number in relation to each marker (artificial afferent) that might model firing rate. There is some evidence to suggest that afferents posses directional selectivity [75, 271, 275] or irregular shaped receptive fields [276], however these aspects are beyond the scope of the present work.

At this stage, we do not perform any manual feature extraction from our audio waveform: to determine the salient features within the raw signal is one of the aims of this study.

4.5.3 Results

Results for the smooth tip are shown in Figs. 4.6 and 4.7 for slow and fast presses respectively. For the fingerprint tip, the same results are shown in Figs. 4.8 and 4.9 respectively.

Each figure shows response vs. time plots for each tactile channel. Bellow these plots processed tactile images are displayed: snapshots of SA-I and FA firings (depicted by opacity) in their spatial arrangements are shown, along with the raw images from which they are derived. The time instances of tactile images, indicated by black lines between Figs. C and D, were chosen to display distinct phases of the stimulation process: free space, press, stationary at bottom of trajectory and retraction.

4.5.3.1 SA-I afferents

In all cases, there was considerable variation in SA-I afferent firing over the course of each press. SA-I firing was greater, relative to initial values, when the TacTip was fully compressed. As expected, the response was sustained with minimal variation until pressure was released. This characteristic was designed for the purpose of mimicking the slow adaptation of SA-I afferents. There was a short time delay after the robot initially moved and the SA-I afferents started to respond since the TacTip must traverse a small gap before contacting the flat surface. We also see that after the robot stopped moving at the bottom of the tap SA-I afferent firing continued to increase for ~0.05 s, likewise SA-I started to fall ~0.05 s after the robot began retracting. We hypothesise that this delay was caused by a small time-lag between sending and executing commands on the robot.

During the release phase, response fell towards zero, however, the response curve is not symmetric; i.e., time taken for SA-I firing to return to zero from maximum firing in the release phase was longer than that of SA-I firing to reach maximum in the press phase. There was a gradual decay towards zero after the TacTip lost contact with the surface. This indicates hysteresis (memory effects) within the TacTip's gel or skin as it takes time to 'relax' to its at-rest form.

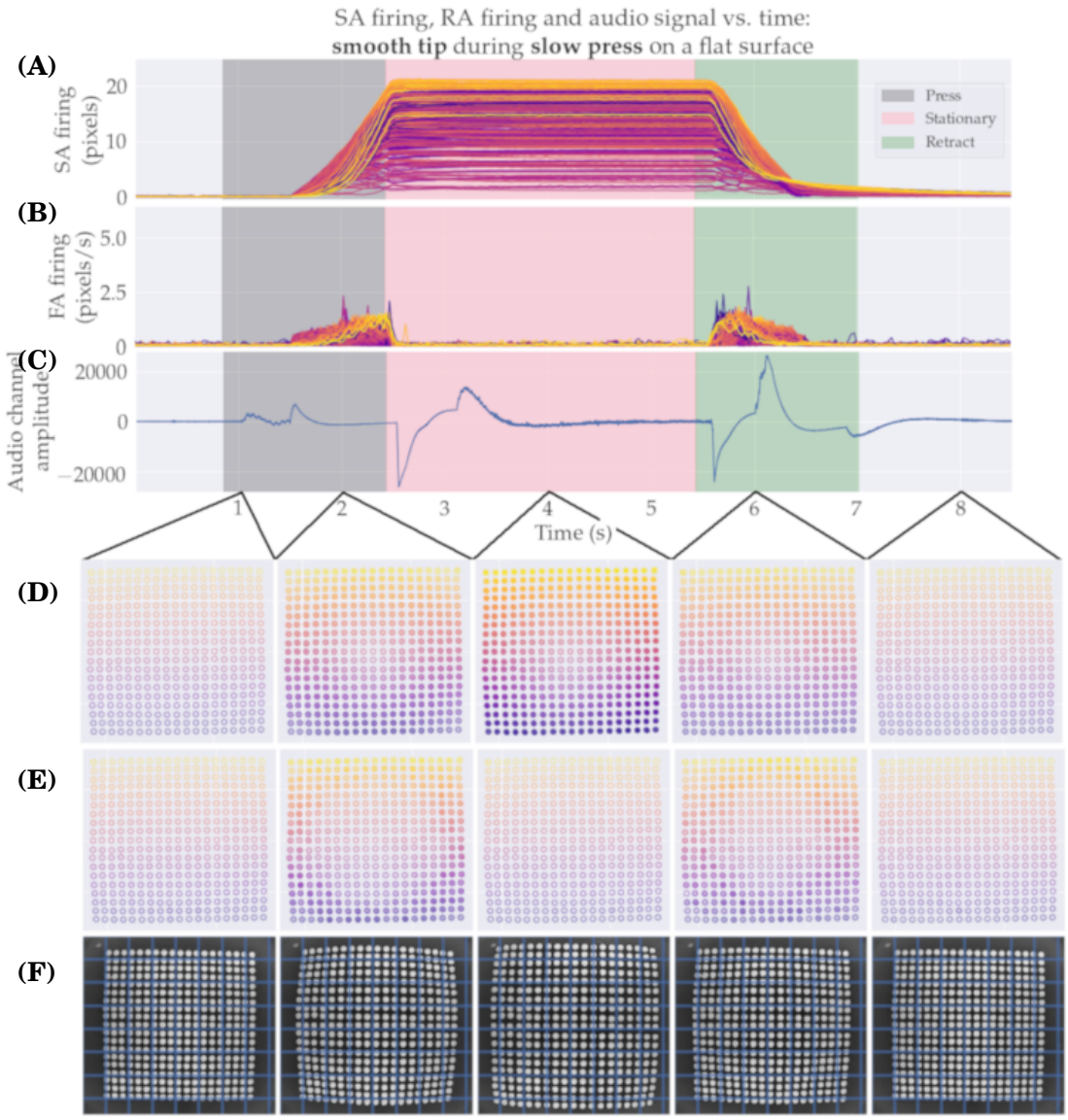

Figure 4.6: SA, FA and vibrational signals vs time collected with the **smooth** tip during a press on a flat surface moving at $\mathbf{3\,mms^{-1}}$, **A**, **B** and **C** respectively. **D** and **E** show the spatial arrangement of SA-I and FA afferents and their respective firing (depicted by opacity of each colour) at times indicated on the axis of **C**. The corresponding raw tactile images are shown in **F**. The colour of each afferent is consistent across **A**, **B**, **D** and **E**

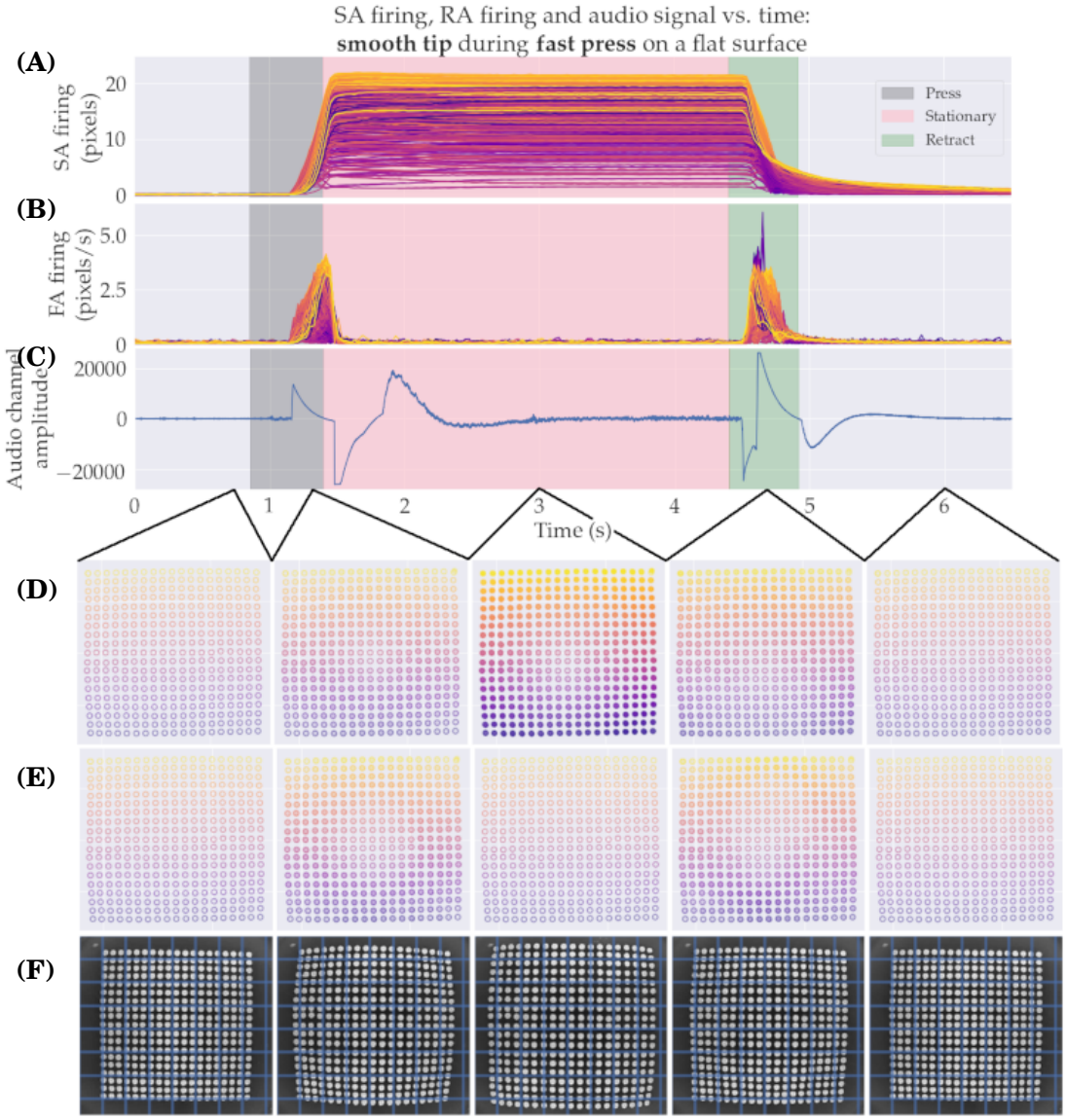

Figure 4.7: SA, FA and vibrational signals vs time collected with the **smooth** tip during a press on a flat surface moving at $\mathbf{10\,mms^{-1}}$, **A**, **B** and **C** respectively. **D** and **E** show the spatial arrangement of SA-I and FA afferents and their respective firing (depicted by opacity of each colour) at times indicated on the axis of **C**. The corresponding raw tactile images are shown in **F**. The colour of each afferent is consistent across **A**, **B**, **D** and **E**

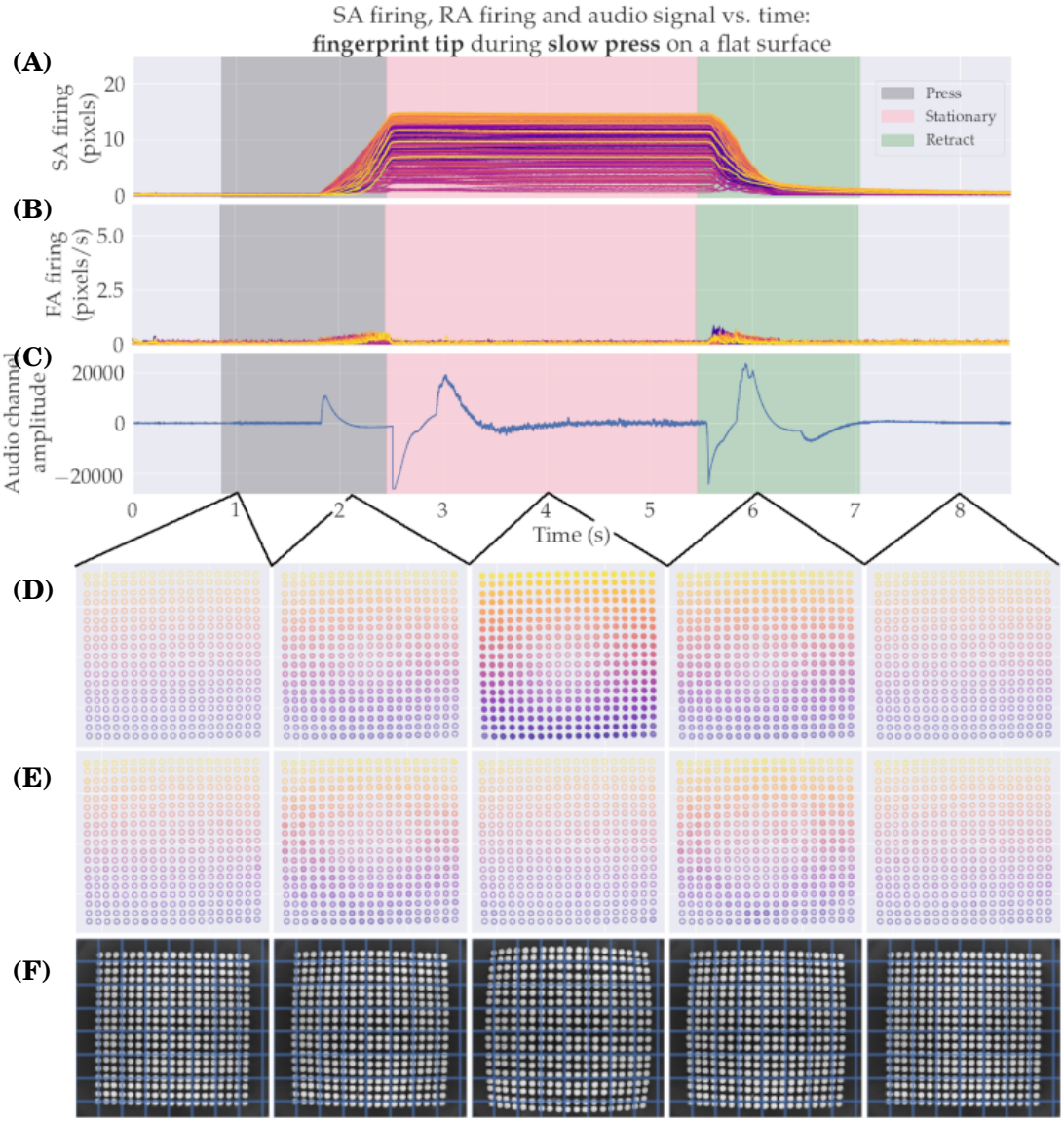

Figure 4.8: SA, FA and vibrational signals vs time collected with the **fingerprint** tip during a press on a flat surface moving at $\mathbf{3\,mms^{-1}}$, **A**, **B** and **C** respectively. **D** and **E** show the spatial arrangement of SA-I and FA afferents and their respective firing (depicted by opacity of each colour) at times indicated on the axis of **C**. The corresponding raw tactile images are shown in **F**. The colour of each afferent is consistent across **A**, **B**, **D** and **E**

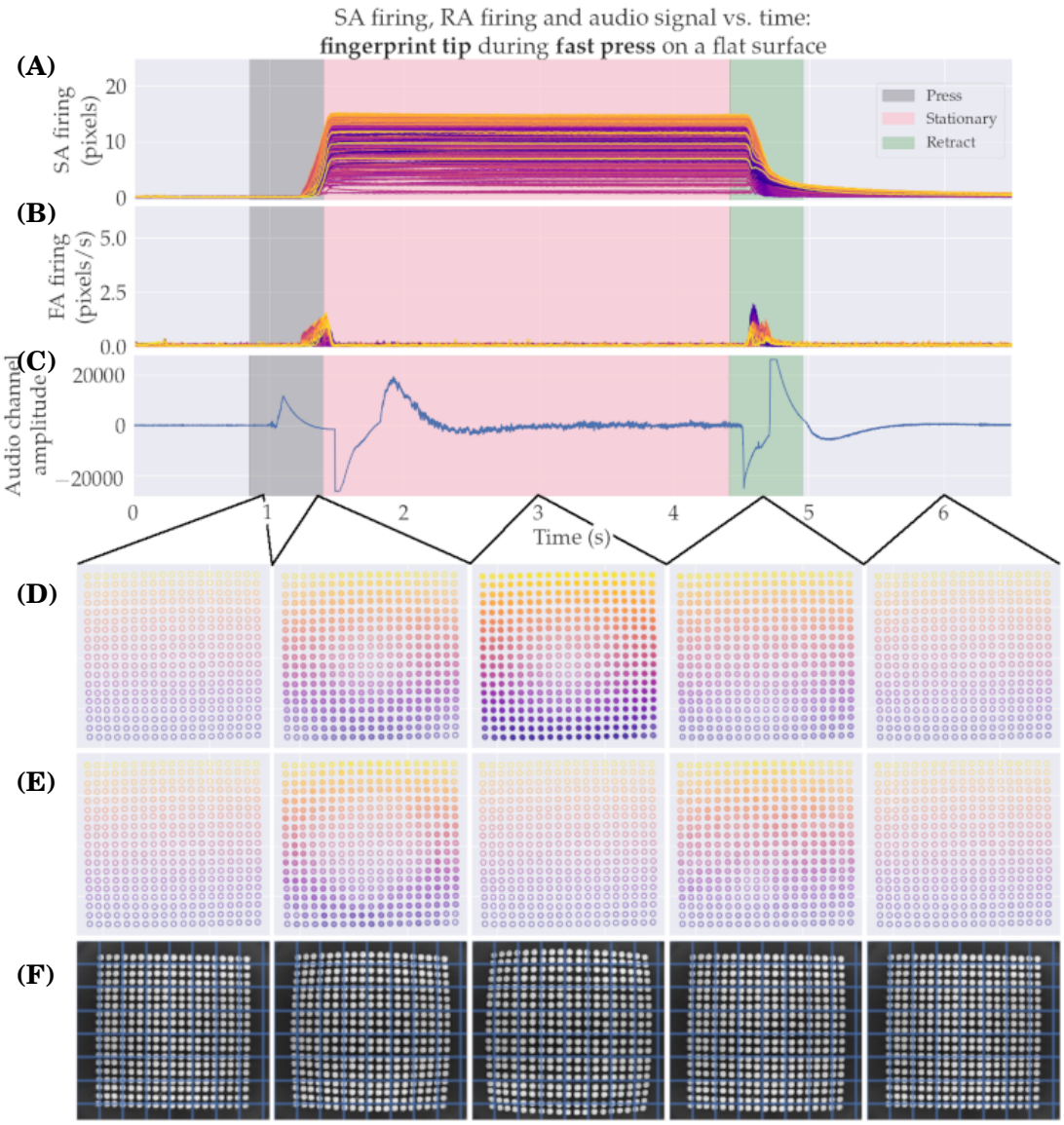

Figure 4.9: SA, FA and vibrational signals vs time collected with the **fingerprint** tip during a press on a flat surface moving at $\mathbf{10\,mms^{-1}}$, **A**, **B** and **C** respectively. **D** and **E** show the spatial arrangement of SA-I and FA afferents and their respective firing (depicted by opacity of each colour) at times indicated on the axis of **C**. The corresponding raw tactile images are shown in **F**. The colour of each afferent is consistent across **A**, **B**, **D** and **E**

The spatial arrangement of SA-I firing was similar in all cases. We observe a central region, which is approximately circular (Figs. 4.6B, 4.7B, 4.8B, 4.9B), where minimal firing occurs, even at points of peak amplitude, as seen in the relative SA-I response plots above. Radiating away from this central region, afferents tended to exhibit increased firing.

This spatial pattern can be understood by considering a crude physical model of the interaction between the TacTip's skin and the stimulus which we depict in Fig. 4.10: When the tip is pressed

onto a flat surface, a circular area of the tip compresses and conforms to the surface. The size of this area correlates to the amount of force exerted on the tip. This causes some pins to spread outwards in the 2D plane viewed by the camera. Individual artificial SA-I afferent firing is a measure of lateral displacement and therefore individual artificial SA-I afferent firing encodes local shear magnitude. Under these conditions, the amount of shear experienced by each marker is roughly dependent on how much the angle of the skin directly below it is affected. In theory, this is none for the central afferent and increasingly more for afferents further towards the edge. Figs. 8A and 8B show the tip pre and post-stimulation respectively. Notice how marker displacement increases for markers further from the centre.

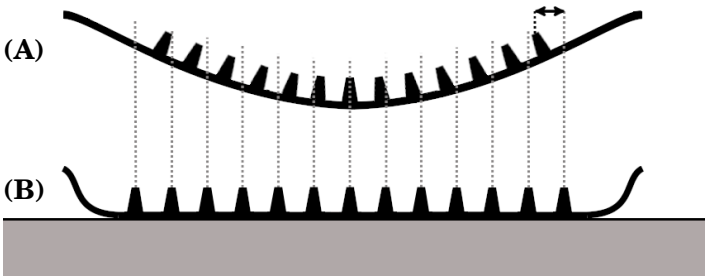

Figure 4.10: Simplified diagram of TacTip deformation when pressed onto flat plate. **A**: small portion of tip at rest; i.e., no deformation. **B**: the same portion of the tip when pressed against a flat surface.

Moving the robot at different speeds had a predictable affect on the SA-I response curves: in both the smooth and fingerprint tip, the gradients during pressing and release phases were increased when the robot moved faster.

The smooth tip exhibited larger amplitudes in SA-I response compared with the fingerprint variant. Whilst possible effects of the fingerprint on SA-I firing under these circumstances cannot be ruled out, it is important to acknowledge that inconsistencies between the manufacturing procedure of each tip are also very likely to have an effect on the differences observed. In particular, the smooth tip was moderately more filled than the fingerprint version. This may have resulted in higher amplitudes because the smooth tip has a a larger 'bulge' which simply means that for a given robot position the the TacTip is compressed slightly further into the stimulus.

4.5.3.2 FA afferents

In contrast to SA-I firing, the FA channel exhibited distinct transience. In all cases, FAs only fired during the dynamic phases of the robot's cycle. Importantly, the response of each artificial afferent behaved as designed, to mimic the adaptation of natural FA afferents, by quickly attenuating the response as the stimulus becomes sustained. As with SA-I firing, we observe short time delays between robot dynamics and expected changes in FA firing which we expect was a product of the same time lag in robot control. It appears that FA firing is more noisy than that of SAs, however, we expect this apparent noise increase is an effect of the reduced numerical amplitude in the derivative signal and thus represents a reduced signal-to-noise ratio.

As with SA-I response, at peak amplitudes, FA firing also exhibited approximate radial symmetry where firing is low in the central region and increased outwards. In this example, dynamics are relatively simple: each marker moves in one direction during the press phase and in the opposite direction during the release phase. SA-I response can be computed as the integral of FA firing with respect to time. Thus, in this particular example it is unsurprising that the spatial arrangements of SA-I and FA firing are similar. In more dynamically complicated scenarios, however, where marker velocities and displacements are less predictable; e.g., sliding, stick-slip or vibration, it is reasonable to assume that the spatial arrangement of FA firing could be different to that of SA-I at any given instance.

The consequence of varying speed had a predictable effect on FA firing: The peak values in FA response were considerably larger when the robot moved faster.

As with SA-I response, we suggest that no meaningful conclusion regarding the fingerprint's effect on FA firing can be drawn from these results due to inconsistencies in the manufacturing process.

4.5.3.3 Vibrational channel

As with the FA response, the vibrational channel output exhibited transience: non-zero amplitudes only occurred in or shortly after phases where the robot was moving and the TacTip is in contact with the stimulus. For example, consider Fig. 4.9: press and release phases are broadly characterised by a sharp rise or fall in amplitude on their respective onsets (~1 and ~4.5 s respectively), followed by some non-linear decay towards zero before a sharp fall or rise in amplitude to their peak negative and positive values respectively (~1.5 s and ~4.75 s respectively). Both phases appear to end with a signal that characterises a near-to critically damped system (to ~2 s and ~5 s respectively): we hypothesise that peak negative amplitudes were created when the robot stopped at the bottom of the press phase and peak positive amplitudes were created when the sensor lost contact with the stimulus. After these peaks, the signal decayed as it oscillated about zero amplitude for roughly one period.

The effect of robot speed on the vibrational channel was less distinct than its effect on SA-I and FA firing. Predictably, the periods in which the vibrational channel responded with non-zero output during press and release phases was shortened from ~2 s to ~1.5 s when the robot moved faster, although this was not by the same factor as the relative speed change ($\frac{3}{10}$). This discrepancy is likely due to the fact that in both cases the system experienced oscillation of pressure post movement: the speed at which pressure waves move through the tip is related to physical properties of the gel rather than dynamics of the robot.

Again, as with both SA-Is and FAs, we do not draw conclusions from the difference between vibration signals for the smooth and fingerprint tips due to inconsistencies in the manufacturing process. As stated, this could be due to a different amount of gel within each tip.

4.5.4 Discussion

Artificial SA-I afferents show no adaptation to consistent pressure, whereas artificial FA afferents rapidly adapt as the stimulus becomes sustained. This result suggests that our artificial SA-I and FA afferents have similar characteristics to their natural counterparts; i.e., compared to Fig 2.2. Furthermore, the physical resolution of SA-I and FA afferents appears to provide information about the shape of contact, as is suspected of both natural SA-I and FA afferents [90, 92].

It is believed that natural PC afferents are generally responsive to high frequency vibrations [44], therefore we do not aim for the vibrational channel to be utilised in the types of stimulation observed here (low temporal frequency). However, it is clear from these results that, as with the artificial FA afferents, the vibrational is responsive only to dynamic touch, suggesting its suitability as a model for natural PC afferents.

4.6 Experiment: Comparison with Physiological Data

To assess the likeness of our model to its natural counterparts, we compare recordings of afferent response from neurophysiological experiments to recordings made with the TacTip under similar conditions.

4.6.1 Neurophysiological Experiment

Phillips and Johnson (1981) [112] present an influential study, in which response of SA-I and FA primary afferents innervating monkey's fingerpads are recorded whilst performing vertical indentations across a series a gratings with aperiodically varying gap and ridge width. The study was a follow-up to a psychophysical paper by the same two authors [43], in which they found that human performance in grating orientation, gap detection and letter recognition tasks, could be accounted for by information contained in spatial modulation of SA-I or FA afferent discharge based solely on their innervation densities within the human fingerpad. The purpose of [112] was to provide evidence, for or against, a spatial code for these types of task, by ascertaining the degree to which primary afferent firing is modulated under similar spatial periods.

The stimuli used by Phillips and Johnson were a set of seven acrylic rectangular waveforms (gratings). Within each grating, ridges at the left and right extremes had variable widths, the internal ridge width remained constant whilst the gaps between ridges varied monotonically in the direction perpendicular to ridge/gap axis. An example of a stimulus cross section is shown in Fig 4.11. The average total length of these gratings was ~24 mm, the width of ridges ranged between 0.5 mm and 5 mm and the number of gaps per grating ranged between 8 and 2. Recordings of 9 SA-I and 5 FA afferents innervating the central region of the macaque monkey's distal fingerpad were recorded whilst indenting a stimulus into the fingerpad to a depth of 1 mm. Between each indentation, the stimulus was stepped 200 μm across the most sensitive region of each afferent. The total period of indentation per step was 1 s. PC afferents were not used as they

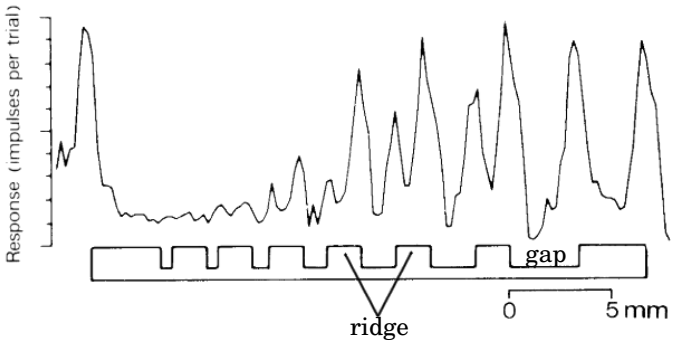

Figure 4.11: Spatial-response profiles (SRPs) of discharge from a single monkey SA-I afferent when indented with at distinct locations across two separate aperiodic gratings (from Phillips and Johnson (1981) [112]).

do not have the required innervation density, nor the receptive field size, to provide the spatial resolution reported in [43].

Results in [112] were displayed as spatial-response profiles (SRPs), shown in Fig. 4.11. As described by Phillips and Johnson (1981): "Each spatial-response profile summarizes the responses of one afferent to many indentations by one grating; between indentations the grating was shifted horizontally by 200 μm along an axis through the centre of the afferent's receptive field and lying at right angles to the grating bars. The response measure plotted against horizontal location is the total impulse count during the 1 s indentation period." [112]. Similar to the liberal interpretation of SEPs [91] (see Section 2.1.3.3 of literature review), an SRP might be considered to approximate a spatial neural image of the stimulus as it would appear distributed across a population of neurons. The key result, demonstrated by these plots, was that when the space between ridges was reduced below a certain distance (~1 mm), the response profile in SA-I or FA did not exhibit enough detail to resolve the stimulus shape. It is likely that this effect is a consequence of the size of the individual afferent's receptive field which may span multiple ridges and/or a lack of compliance in the tissue of the fingertip which can act as a low-pass filter, attenuating high spatial frequencies. Either way, it demonstrates a similar limit on the spatial resolution as observed in the preceding paper [43].

Here we execute, as far as possible, this expetiment with the TacTip. We are interested to observe similarities and differences between artificial SA-I and FA afferent response to the natural counterparts and also to understand the limits on resolution presented by single afferent response.

4.6.2 Data Collection

4.6.2.1 Stimuli

In this study, we use a set of seven 3D-printed gratings as stimuli. The stimuli correspond to those used in [112], with ridge and gap widths varying in the same proportions. The scale used in our study, however, is doubled, thus all ridges and gaps are twice as wide. This scaling was chosen since the innervation density of markers in the TacTip (~40 cmm^{-2}) is approximately half that of

Table 4.1: Table lists ridge and gap widths of the aperiodic gratings used.

	Ridges (mm)											
Grating	**left**	**internal**	**right**	**Gaps (left to right, mm)**								**Overall length (mm)**
A	6.0	1.0	6.0	1.0	1.0	1.5	2.0	3.0	4.0	6.0	10.0	47.5
B	6.0	1.5	6.0	1.0	1.0	1.5	2.0	3.0	4.0	6.0	10.0	51.0
C	3.0	2.0	6.0	1.0	1.0	1.5	2.0	3.0	4.0	6.0	10.0	51.5
D	6.0	3.0	6.0	1.0	1.0	1.5	2.0	3.0	4.0	6.0		48.5
E	6.0	4.0	6.0		1.0	1.5	2.0	3.0	4.0	6.0		49.5
F	6.0	6.0	6.0				2.0	3.0	4.0	6.0		45.0
G	10.0	10.0	10.0							6.0	10.0	46.0

SA-I afferents in the human fingertip (see Section 2.1.1.1). Table 4.1 gives a full description of the dimensions of the stimuli used.

4.6.2.2 Collection Procedure

A photograph of the experimental setup is shown in Fig. 4.4, where the 7 stimuli are the gratings described in Table 4.1. The robot makes successive taps onto a stimulus indenting the skin by ~1 mm. Each tap takes ~0.8 s, the robot is stationary at the bottom of a tap for 0.5 s. Between each tap, the robot moves 200 μm in the direction perpendicular to the ridge/gap axis, starting in free space on the left-hand side and moving over the entire stimulus until free space is again reached (left and right as seen in Fig. 4.12). TacTip video data is simultaneously recorded for the entire period of each tap. This process is repeated seven times: once for each stimulus.

4.6.3 Results

As in [112], results are displayed in the form of spatial-response profiles (SRPs). In the case of SA-I and FA artificial afferents, data is collected as described in Section 4.6.2 and responses are acquired from pin positions using the procedure described in Section 4.5.2. SRPs show the mean response at each position, separated by 200 μm, during the course of each indentation.

We review the results in the context of four features:

(i) Effect of individual ridges and edges.

(ii) Effect of neighbouring ridges and edges.

(iii) Variation across afferents.

(iv) Effect of fingerprint.

(i) and *(ii)* are taken directly from [112]. *(iv)* is similar in affect to 'grating orientation' from [112], which is designed to manipulate the orientations of stimulus edges relative to papillary ridges of the monkey's fingerpad.

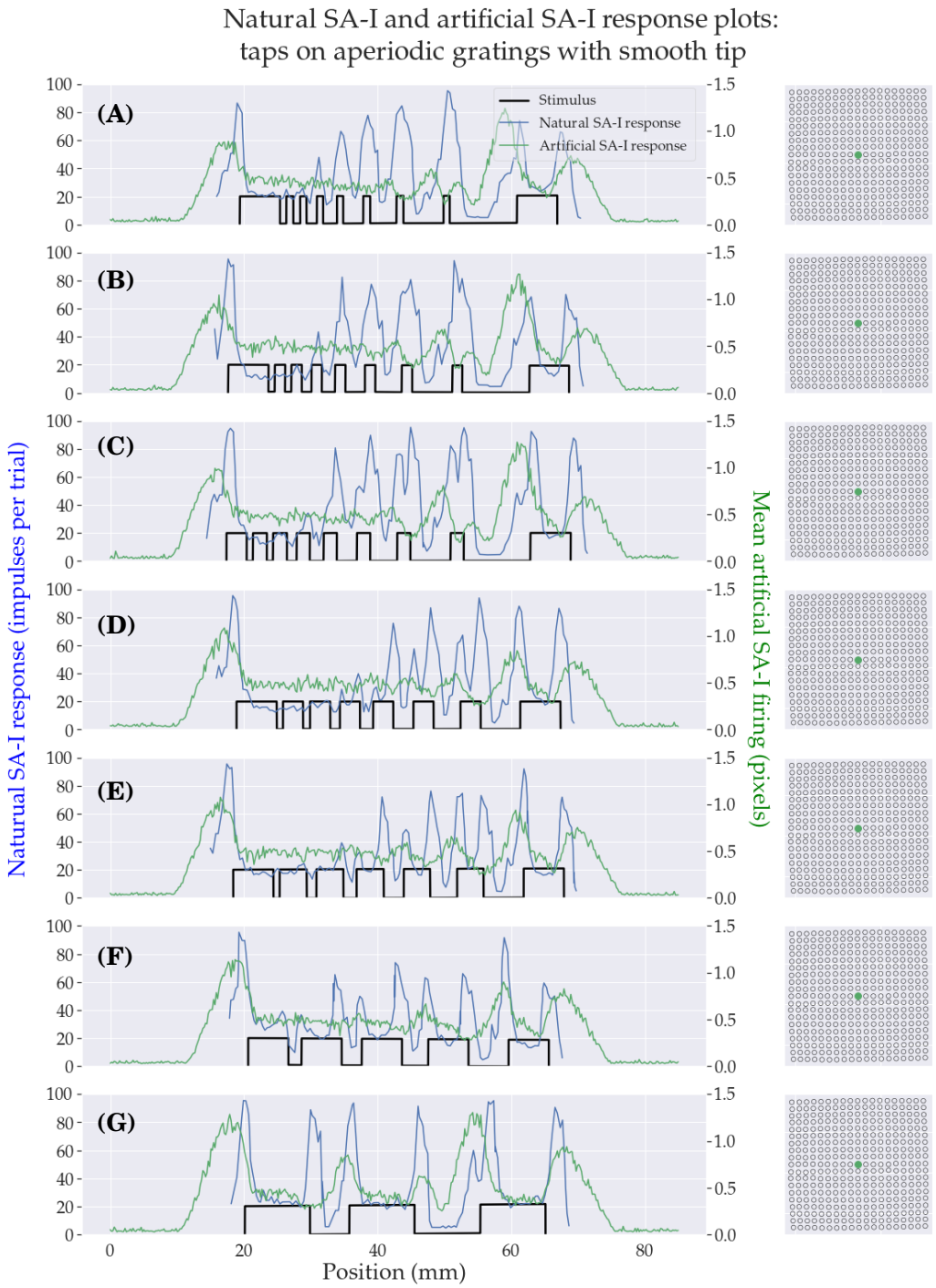

Figure 4.12: *Left*: SRPs for a single monkey SA-I afferent collected on seven gratings described in [112] (blue) and SRPs for the central artificial SA-I afferent of the smooth tip (green) collected on gratings (black). *Right*: position of central artificial SA-I afferent (green).

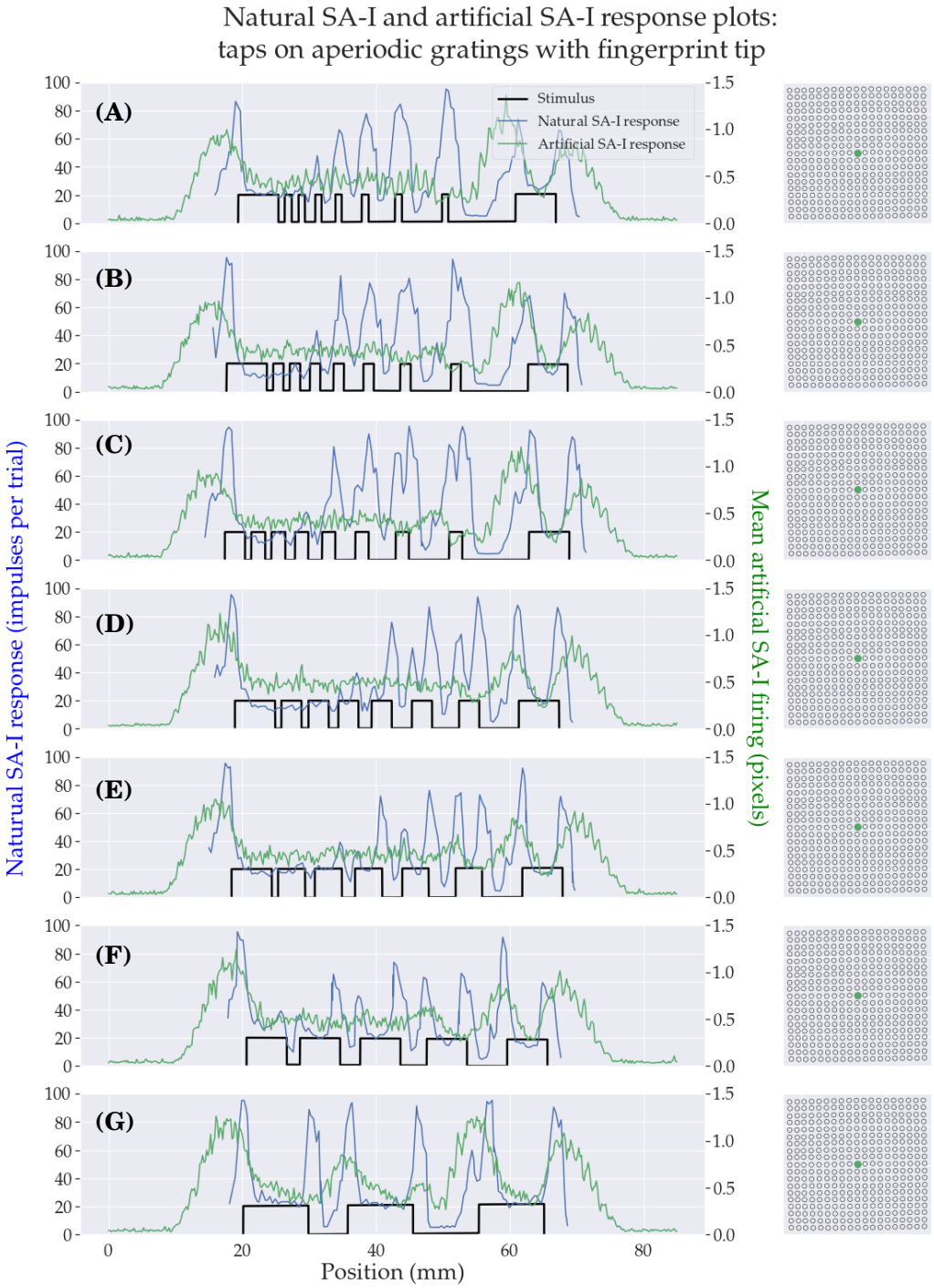

Figure 4.13: *Left*: SRPs for a single monkey SA-I afferent collected on seven gratings described in [112] (blue) and SRPs for the central artificial SA-I afferent of the fingerprint tip (green) collected on gratings (black). *Right*: position of central artificial SA-I afferent (green).

4.6.3.1 SA

Figs. 4.12 and 4.13 show SRPs for a single artificial SA-I afferent (green), smooth and fingerprint tips respectively, on the 7 aperiodic gratings described in Table 4.1 and equivalent SRPs of a single monkey SA-I afferent (blue) collected on similar gratings (described in [112]).

Effects of Individual Ridges and Edges:
Phillips et al. state that "all of the SA-I afferents yielded a much larger response to an isolated narrow bar or an edge facing a large gap than to a flat surface" [112]. Perhaps the most striking resemblance between the artificial SA-I afferents and their natural counterpart is their preferential sensitivity to 'edges facing a large gap'. This is most clearly seen in Figs. 4.12G and 4.13G, where the response of natural and artificial SA-I afferents were attenuated by the 3 flat surfaces and, in general, amplified by the 6 edges. This characteristic is also readily seen on the approach and departure of each grating, where gratings exhibit rising and falling edges before and after large gaps respectively. SRPs for artificial SA-I afferents, however, do not exhibit the same sensitivity to isolated bars as monkey SA-I afferents. In fact, it may be accurate to conclude that preferential sensitivity for artificial afferents was limited to edges since amplification, if any, always coincided with the rising or falling edge of a ridge.

The observed preferential sensitivity of artificial SA-I afferents to edges vs. flat surfaces can be explained by the model described in Section 4.5.3.1. In our model, SA-I afferents respond only to lateral displacement as viewed by the camera. As described in Section 4.5.3.1, the central afferent, in theory, does not respond to vertical forces with flat surfaces (see Fig. 4.10); e.g., ridges. In contrast, when the stimulus is an edge, the marker is likely to move laterally owing to the levering effect of the associated pin, thus the afferent responds. This is depicted in Fig. 4.14 **A** and **B**.

Effect of Neighbouring Ridges:
Phillips et al. on the effects of neighbouring ridges: "it appears that the effects of surrounding stimuli are negligible at 3.0 mm. As the bars are spaced more closely than 3.0 mm the heights of the associated response peaks are diminished". This comment refers to scales at which monkey afferent response is able/unable to distinguish separate stimuli and is an important concept in determining the ability to discriminate these scales using a spatial code of primary afferent firing. Artificial SA-I firing did not exhibit sensitivity to individual narrow ridges, as did its natural counterpart, however, the enhanced sensitivity to edges appears to exhibit similar modulation with stimulus scale. For example, upon examining Figs. 4.12A,4.12B and 4.12C, we observe that the second ridge from the right elicited amplified edge response on both the rising and falling edges. This ridge is separated by 6 mm from the preceding in all three cases. As we consider further left in each figure, where gap width decreases, the degree of edge enhancement rapidly attenuates. As gaps become closer than ~4 mm artificial SA-I afferents appear to consistently responded to sets of individual ridges as though they comprise a single flat surface for both the

smooth and fingerprint tips.

The observed effect of neighbouring ridges is a consequence of the physical properties of the TacTip skin. Notice how the skin, depicted in Fig. 4.14 **B**, does not perfectly conform to the stimulus edge. Essentially, the TacTip can only deform a certain amount owing to finite flexibility. A consequence of this is that when a stimulus edge is closely neighboured by another edge, the tip can deform even less. This is depicted in Fig. 4.14. Notice how the marker displacement in Fig. 4.14 **D** is less than that of Fig. 4.14 **C**. The closer each stimulus ridge becomes, the more the entire stimulus 'feels' like a flat surface and, therefore, the less the central afferent will fire. This is the essence of why stimulus edges are less tangible when closely neighboured by other edges. In this way, the skin acts as a low-pass filter, attenuating high spatial frequencies.

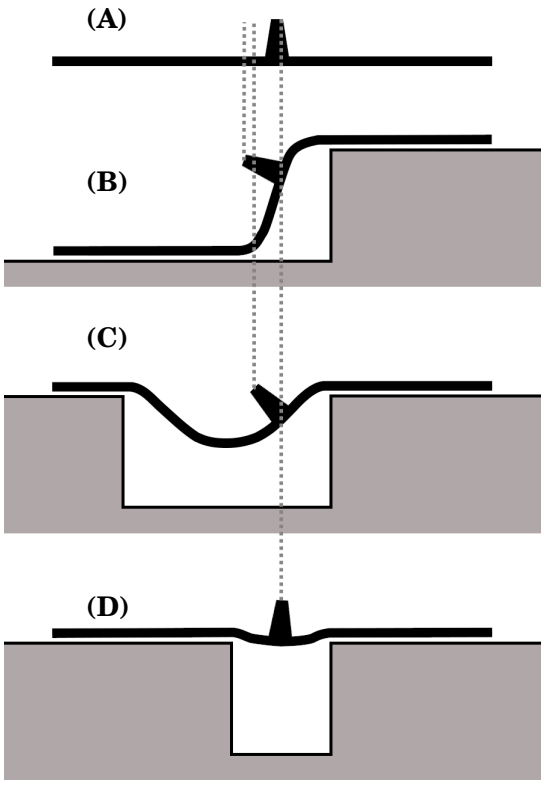

Figure 4.14: Simplified diagram of TacTip central afferent when pressed onto edge stimulus. **A**: small portion of tip at rest; i.e., no deformation. **B**-**D**: the same portion of the tip when pressed on an edge neighbouring an increasingly smaller gap.

Variation Across Afferents:

Phillips et al. report edge enhancement ratios of different SA-I afferents (ratio of amplitude at edges to that at flat surfaces) ranging from 1.5 to 20 with typical values between 2.0 and 6.0. Whilst the edge enhancement ratio was less for the case of artificial SA-I afferents (average of ~2.0 for the central artificial SA-I afferent of both smooth and fingerprint tips: Figs. 4.12 and 4.13), we certainly observe a significant range for this feature: Artificial SA-I firing in SRPs in Figs. 4.12 and Figs. 4.13 were derived from the central marker which showed the greatest edge enhancement. For comparison we show the SRPs for another artificial afferent (selected at random), on the smooth and fingerprint tips, collected on grating **A** in Fig. 4.15. (SRPs for a further 5 artificial SA-I afferents are displayed in Appendix A, Section A.2.1, Fig. A.6). Notice how these afferents showed very little, if any, enhanced sensitivity to edges. Interestingly, some off-centre afferents (e.g., Fig. 4.15) showed an increased intensive response, that is the total firing over the entire stimulus, when compared to the central afferent. Others (see Appendix A, Fig. A.6) fired less.

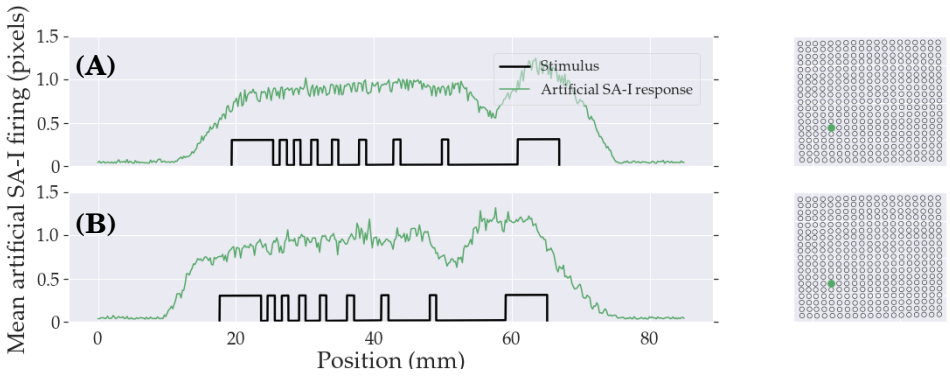

Figure 4.15: *Left*: SRPs for a non-centrally located artificial SA-I afferent of the smooth tip (**A**) and fingerprint tip (**B**) (green) collected on grating **A** (black). *Right*: position of selected artificial SA-I afferent (green).

Phillips et al. assign the variation in afferent's response profiles to differences in receptive field properties: "There were consistent differences between the responses of single afferents, which must be attributed to differences in receptive-field properties." In contrast, we believe that each artificial afferent is unlikely to exhibit drastically different shaped receptive fields. In fact, what is more likely to have caused the variation in SRPs of artificial SA-I afferents is the spatial arrangement of markers from which artificial SA-I response is derived, in combination with the topographic variation in TacTip's skin height in the position of each afferent's receptive field. For example, we have shown how only the central artificial afferent exhibits no response when pressed against a flat surface; i.e., a ridge (Fig. 4.10).

Effect of Fingerprint:

Phillips et al. consider the effect of grating orientation relative to the skin ridges on monkey fingerpads [112]. Rather than the characteristic 'whorl' of human fingerprints, monkey's papillary ridge pattern tends to form parallel rows in line with the finger's primary axis. Phillips et al. observed that whilst grating orientation had no effect when the receptive field was above a flat surface, the edge enhancement ratio was increased when stimulus ridge aligned with papillary ridges.

An artificial fingerprint had little effect on edge enhancement of artificial SA-I afferents in comparison to their natural SA-I counterparts. The addition of a fingerprint, however, appears to have increased apparent noise into the response profiles for the centrally located afferent (Fig. 4.13). One explanation for this might be an interaction between each individual exterior nodule and stimulus ridge edge: it is possible that a small degree of amplification occurs when the fingerprint nodule 'catches' a stimulus edge, essentially levering the pin. In this case, the modulation frequency is likely to occur on the same scale as fingerprint pattern (3 mm) which is on the scale of what is seen in Fig. 4.13.

In both smooth and fingerprint tips we also observe a systematic preferential sensitivity to

left edges. This is particularly obvious in Figs. 4.12G and 4.13G, where there is no amplified response on the right hand edge of the left most ridge and to a lesser degree in Figs. 4.12 and 4.13 on the left and right hand edges of the right most ridge. Considering a perfectly aligned experimental setup and a TacTip with symmetric physical properties, it does not seem possible for the system to exhibit such systematic preferences to one type of edge or another. Given that the same effect seems to be present in both tip types suggests that this effect is a consequence of the experimental setup. For instance, the vertical tap may have be slightly off axis or the test bed may have been not perfectly horizontal.

4.6.3.2 FA

Phillips et al. display response profiles for five FA afferents on aperiodic grating **A** [112] (reprinted in Appendix A, Section A.2.1, Fig. A.8) but do not show FA SRPs on the remaining 6 gratings. For completeness, we show SRPs for selected artificial FA afferents on all 7 gratings of the smooth and fingerprint tips in Figs. 4.16 and 4.17 respectively.

Effect of Ridges and Edges:
Phillips et al. observed that FAs were modulated by the largest (5 mm) gaps on their stimuli, but gaps of 0.5 - 1.0 mm were not represented in any of the response profiles . Additionally, they saw no evidence of edge enhancement [112].

We observe preferentially sensitivity towards edges for artificial FA afferents on both the smooth and fingerprint tips. This distinguishes TacTip's artificial FA afferents from their natural counterpart. However, the edge enhancement ratio for artificial FAs was considerably lower than that of artificial SA-I afferents. One might also conclude that the spatial resolutions exhibited by central artificial FA afferents, in both tips, was marginally lower than that of SA-I. This is perhaps most noticeable when comparing SRPs across afferent types from stimuli **F** and **G**.

Phillips et al. consider two explanations for the difference in SRPs between natural SA-I and FA afferents:

(i) FAs only fire during the initial period of stimulation in each indentation due to their adaptation rates. If skin on the monkey's fingerpads tend to first deform uniformly across the whole pad and then, subsequent to FA adaptation, the skin "flows up around the gratings", one might expect the characteristic structured response from SA-I afferents and a comparatively unstructured response from FAs.

(ii) "An intrinsic difference in the spatial organization of their receptor mechanisms."

Phillips et al. discard theory *(i)* since they observed that SA-I afferents exhibit a structured response even during the first 100 ms of each indentation and therefore conclude that the spatial arrangement of receptive fields is the cause of the difference in SRPs of natural SA-I and FA afferents.

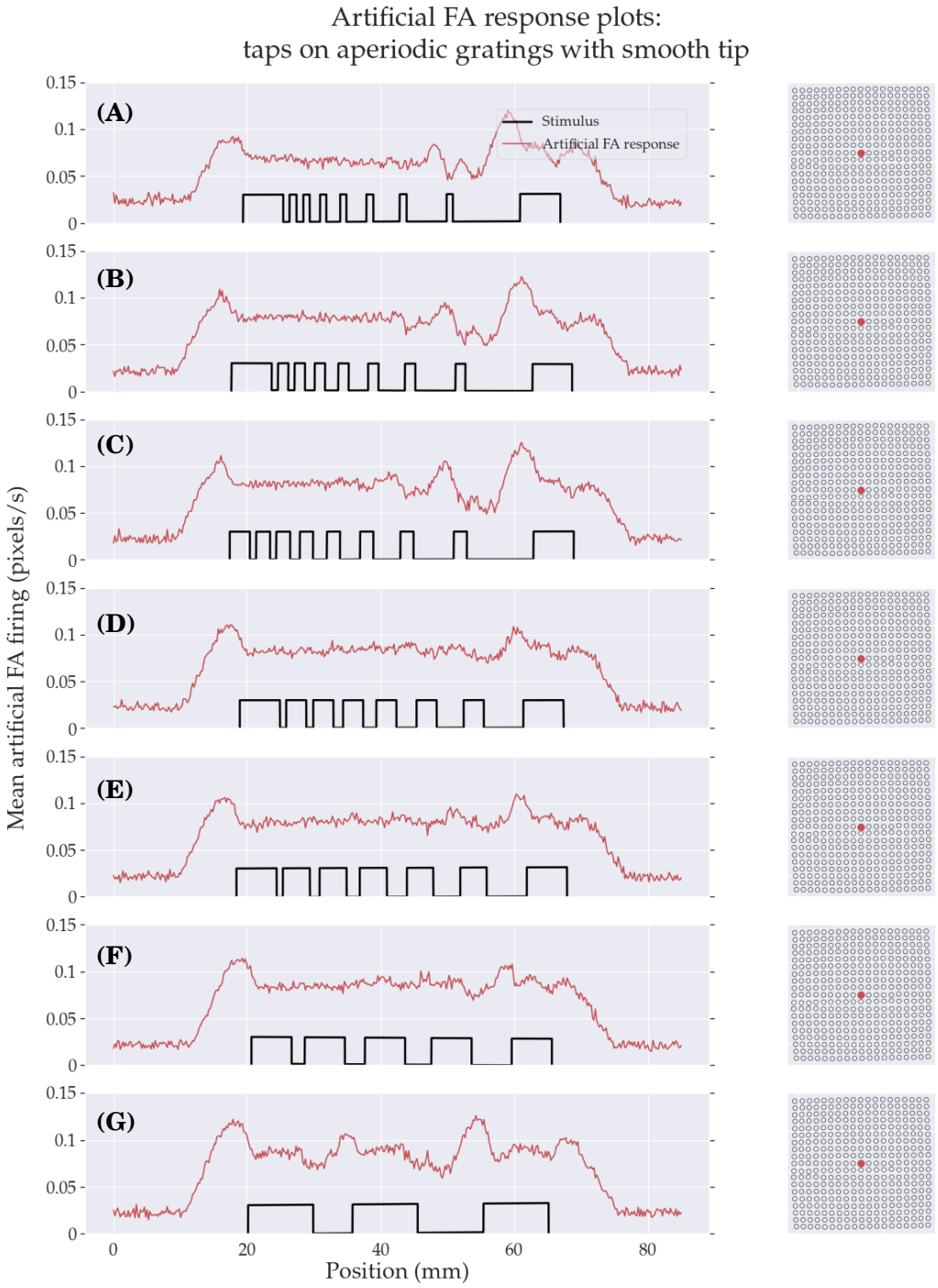

Figure 4.16: *Left*: SRPs for the central artificial FA afferent of the smooth tip (red) collected on gratings (black). *Right*: position of central artificial FA (red).

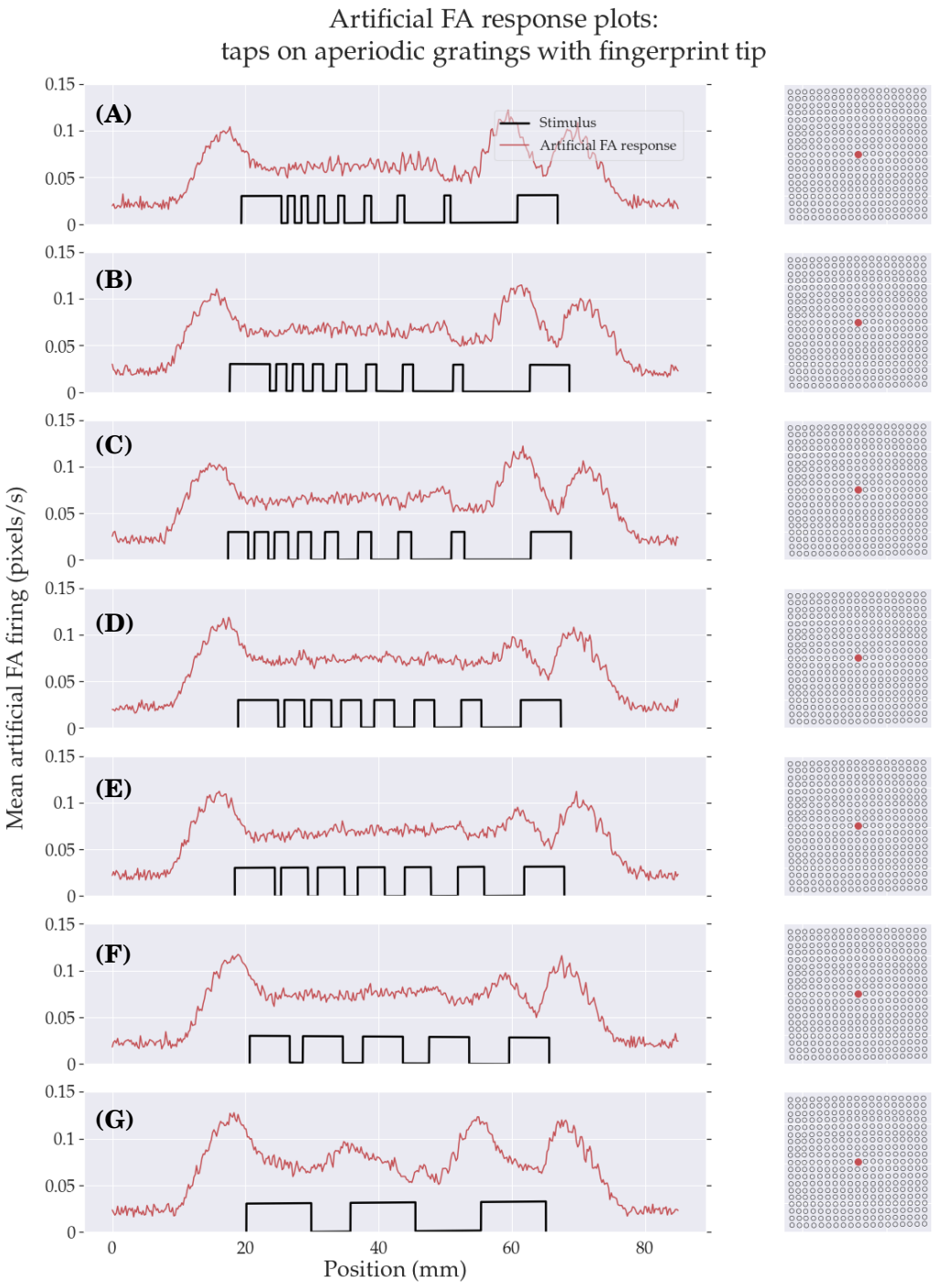

Figure 4.17: *Left*: SRPs for the central artificial FA afferent of the fingerprint tip (red) collected on gratings (black). *Right*: position of central artificial FA (red).

In the case of the TacTip, the spatial arrangement of receptive fields of SA-I and FA afferents are identical since they are both derived from the arrangement of markers. Therefore, we must discard theory *(ii)* as an explanation for the difference in artificial SRPs. Contrastingly, theory *(i)* does seem rather compelling in our case; we have previously observed hysteresis within

the TacTip (Section 4.5.3.1) the physical cause of which might also result in the sort of phasic deformation hypothesised by Phillips et al. across the plane of the tip.

Variation Across Afferents:

Phillips et al. observed that all measured FA afferents shared features such as no preferential sensitivity to edges and poor spatial resolution. Of the five FA afferents found in [112], the only discernable feature in their respective SRPs is the intensity; i.e., the total amount of firing across the entire grating (see Appendix A, Fig. A.8). Peak amplitudes ranged from ~8 to ~20 impulses per trail.

Fig. 4.18 shows SRPs for an off-centre artificial FA afferent on smooth and fingerprint tips collected on grating **A** (this afferent was derived from the same marker as used in Fig. 4.15). (SRPs for a further 5 artificial FA afferents are displayed in Appendix A, Section A.2.1, Fig. A.7). As with SA-I variants, off-centre artificial FA afferents showed very little, if any, edge enhancement. We also observe reduced resolution, where all but the right most ridge were 'felt' as a single feature. We believe the variation observed in SRPs of artificial FA afferents stems from the same physical behaviour offered as an explanation for the variation on artificial SA-I response profiles. That is, the variation in alignment of the force exerted by the stimulus on the TacTip, which is caused by the slight dome of the TacTip's skin.

Effect of Fingerprint:

We do not observe any significant effect of an artificial fingerprint on FA response. This implies that whilst marker deflection is appreciably affected by the addition of fingerprint, marker velocities, comparatively, are not.

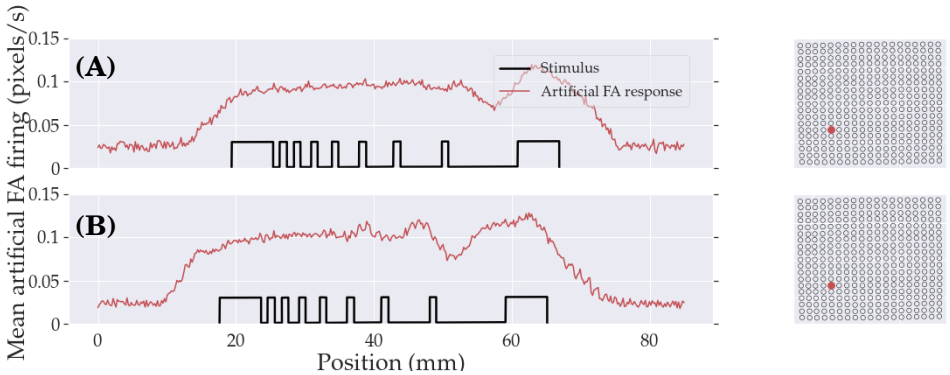

Figure 4.18: *Left*: SRPs for a non-centrally located artificial FA afferent of the smooth tip (**A**) and fingerprint tip (**B**) (red) collected on grating **A** (black). *Right*: position of selected artificial FA afferent (red).

4.6.4 Discussion

Phillips et al. showed that monkey SA-I and FA response profiles were very different. FA firing lacked sufficient detail required to resolve stimuli spaced closer than 2-3 mm [112]. SA-I responses were comparatively highly detailed. These afferents were preferentially sensitive to narrow ridges and edges versus flat surfaces. SAs showed evidence of spatial modulation on the scales that are reliably discriminated in psychophysical experiments [43]. This could provide evidence for a spatial coding in peripheral neural discharge of SA-I afferents for spatial discrimination tasks such as those examined in [43].

Our modelled SA-I and FA afferents exhibited similarities with their natural counterparts. Notably, the preferential sensitivity of SA-I afferents to edges and the modulation of SA-I and FA response with gap width. Furthermore, as with natural afferents, the degree of edge enhancement was reduced with FAs compared with SA-I, so too was the resolution.

We found that artificial SA-I afferents transduce less detail compared with their natural counterparts. For example, in some cases individual ridges were detected by natural SA-I afferents when spaced as closely as 0.75 mm. Contrastingly, the central artificial SA-I afferent for smooth and fingerprint tips were unable to detect ridges spaced closer than ~4 mm apart.

The difference between FA and SA-I response was less pronounced in the case of artificial afferents. Phillips et al. were able to rule out natural FA afferents as a candidate channel for mediating spatial code of stimulus shape based on the detail seen in SRPs. However, in the case of our modelled afferents, the resolving power of FAs does not appear to render it appreciably less useful in providing spatial codes.

The models developed to understand SRP results, depicted in Figs. 4.10 and 4.14, imply that each artificial afferent (SA-I and FA) does not respond to normal force; i.e., a force applied perpendicular to the TacTip's skin, directly in-line with the relevant marker and therefore does not encode compressive or normal strain. This is seen in practice with low firing rates of centrally located afferents in Figs. 4.6, 4.7, 4.8 and 4.9. This theory does not hold completely since, under normal forces, non-central markers will move slightly in the camera's view due perspective effects. Nonetheless, the predominant physical aspect resulting in artificial afferent firing is lateral displacement of markers. This likely presents a discrepancy between natural and presented artificial afferents. Whilst there is some evidence to suggest that shear strain provides a particularly robust representation of stimulus spatial properties; e.g., edges, through spatial coding in afferent populations [277], there is no clear consensus as to the primary local physical stimulus properties determining individual afferent response; e.g., strain energy density [278] and stress [279–281] have both been proposed. Notably authors Phillips and Johnson employed continuum mechanics to model skin dynamics under stimulation with identical aperiodic gratings used in this study and in their neurophysiological experiment [112] to show that compressive shear strain matched firing rates of SA-I afferents with remarkable accuracy [282]. Whilst it is likely that natural afferents do not simply encode lateral strain and its derivative as do artifical

SA-I and FA afferents respectively, we have demonstrated that there are clear similarities between their response characteristic to certain types of stimulation which is nonetheless a significant progression towards replicating natural afferents and producing biologically plausible methods of robot tactile perception.

4.7 Experiment: Dynamic Touch

A common hypothesis in biology is that, under certain conditions, dynamic touch is essential in mediating transient information required for specific tasks [57]; e.g., texture discrimination, friction estimation or slip detection. Here we define dynamic touch as a relative motion between the sensor and stimulus; i.e., sliding.

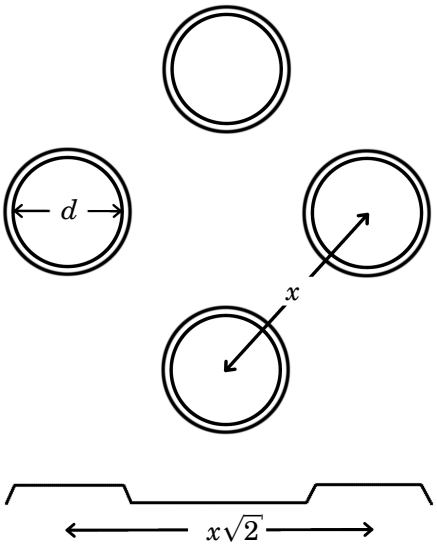

Figure 4.19: Plan and cross-section views of raised bump stimuli showing stimulus dimensions x and d.

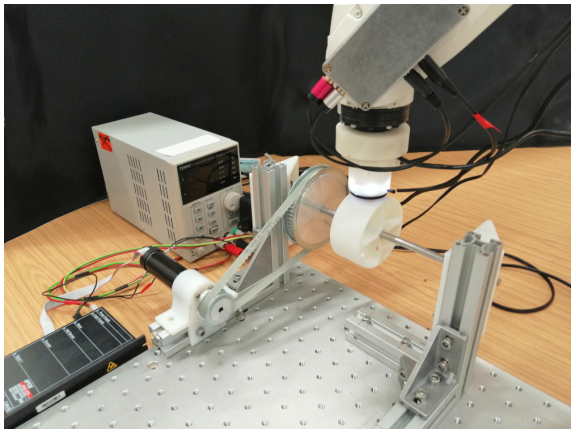

Figure 4.20: Experimental set up for dynamic touch experiments. TacTip is held stationary against a rotating textured drum. The rotating drum is driven by a DC motor and a custom build rig.

In this study we consider dynamic touch and examine the ability of our modelled SA-I, FA and PC channels for mediating salient information relating to textures under sliding motions. Stimuli comprised of periodic raised bumps are used and two textural dimensions are varied: bump separation and bump diameter.

Examining the output from artificial SA-I and FA channels might confirm whether these channels can be implemented for texture discrimination as well as reveal candidate codes; e.g., temporal or spatio-temporal, for objective quantities; i.e., bump spacing and bump diameter. Simultaneously, by examining the nature of vibrations detected by the TacTip's vibrational channel we may understand if cues such as frequency, or intensity relate to these objective

Table 4.2: Complete set of stimuli used. Refer to Fig. 4.19 for dimensinos $\boldsymbol{x}$ and $\boldsymbol{d}$.

Stimulus		Dimensions (mm)
$\mathbf{A_1}$		$x = 2$, $d = 1$
$\mathbf{A_2}$		$x = 2$, $d = 1.5$
$\mathbf{A_3}$		$x = 2$, $d = 2$
$\mathbf{B_1}$		$x = 2.5$, $d = 1$
$\mathbf{B_2}$		$x = 2.5$, $d = 1.5$
$\mathbf{B_3}$		$x = 2.5$, $d = 2$
$\mathbf{C_1}$		$x = 3$, $d = 1$
$\mathbf{C_2}$		$x = 3$, $d = 1.5$
$\mathbf{C_3}$		$x = 3$, $d = 2$
$\mathbf{D_1}$		$x = 3.5$, $d = 1$
$\mathbf{D_2}$		$x = 3.5$, $d = 1.5$
$\mathbf{D_3}$		$x = 3.5$, $d = 2$
$\mathbf{E_1}$		$x = 4$, $d = 1$
$\mathbf{E_2}$		$x = 4$, $d = 1.5$
$\mathbf{E_3}$		$x = 4$, $d = 2$

quantities and, therefore, can be employed in some way for making perceptual judgements about texture.

4.7.1 Data Collection

4.7.1.1 Stimuli

For stimuli we use a set of 15 textured drums. The drums are 3D-printed in a rigid plastic (VeroWhite, Stratasys) using an PolyJet printer (Objet, Stratasys). Each textured drum is patterned with a tetragonal array of raised bumps on its exterior circumference. The bumps are raised 0.5 mm beyond the drum's 80 mm diameter. Both the bump separation, x, and the bump diameter, d, are systematically varied. Fig. 4.19 describes how x and d are defined. Five bump separations, 2, 2.5, 3, 3.5 and 4 mm, are used. At each bump separation three bump diameters, 1, 1.5 and 2 mm, are used.

4.7.1.2 Collection Procedure

The TacTip, mounted as an end-effector to the ABB arm, is held stationary. It is stimulated by a rotating textured drum (80 mm diameter) (see Fig. 4.20). The drum is driven by a 60 W brushed motor which is connected to the drum spindle via a 4.8:1 planetary gear head and a custom built 3:1 belt drive. The spindle is seated between two pillow block bearings and mounted to a Rexroth frame via two 3D-printed brackets. The motor is powered with a 24 V PSU and the velocity is controlled, in closed-loop, via an EPOS2 control unit connected to a PC running consumer software (EPOS studio, Maxon motors). A digital encoder (1024 counts per turn) receives a signal from the control unit and thus drives the motor. The DC-motor, planetary gearhead, control unit and digital encoder are all provided by Maxon motors.

For collecting data a drum is indented ~1 mm into the TacTip's skin by the ABB robot arm. The drum is rotated for 30 s with its circumference moving at

50 mms^{-1} while the TacTip is held stationary. 30 s of TacTip video and audio data are simultaneously recorded. In total we record 30 sets of data: 15 sets (one per stimulus) per tip variant (smooth and fingerprint).

4.7.2 Results

Data is presented in the form of signal (artificial SA-I, artificial FA and vibration) vs time. For each tactile channel, we first show data collected with the smooth tip followed by data collected with the fingerprint tip.

4.7.2.1 SA

Figs. 4.21 and 4.22 show 1.2 s samples of artificial SA-I firing, for smooth and fingerprint tips respectively, when the TacTip has undergone stimulation as described in 4.7.1.

It is difficult to assess the degree to which bump separation or bump diameter produced a systematic change in the SA-I signals for either the smooth or fingerprint tips. However, it might be said that afferents to the right as seen in Figs. 4.21 and 4.22, which were at the trailing edge of contact, generally showed greater responses. This might indicate that purely spatial properties of afferent firing, either instantaneous or accumulated over time, can encode direction of motion and that this information is available in both the smooth and fingerprint variants. We hypothesise that the difference in intensive firing observed across the surface of each tip was a consequence of a low frequency shear force exerted by the stimulus on the tip (here we consider low frequency to be periods of ~0.5 s or more) the mean value of SA-I firing within a ~5 s window, corresponding to ~25 mm of stimulus substrate, can be considered a DC-bias and is related to, not the texture itself, but rather the drum diameter which is likely to exhibited a small variation of ~0.2 mm.

The fingerprint tip shows a significantly increased sampling rate compared with the smooth version. This is intriguing since all hardware aside from the tip itself was the same including the camera. Upon initially noticing this, a number of informal tests were performed, including with different versions of the same camera, which all confirmed that when using the smooth tip, the frame-rate was drastically reduced to ~30 Hz compared with maximum of ~90 Hz achievable with the fingerprint tip. A probable source is of this discrepancy is some sort of compression or optimization taking place upon the camera's hardware. This is difficult to diagnose and fix. Consequently we have been unable to match the frame-rates between the two tips.

It is important to consider aliasing that will inevitably take effect as a consequence of sampling. When the signal frequency, f, is more than the Nyquist frequency, $f_{\text{Nyquist}} = \frac{g}{2}$, where g is the sample frequency, aliasing will take effect: the signal is reconstructed at a lower frequency and therefore will corrupt the measurement. The reconstructed frequency is given by the formula below:

$$f_r = \left| f - g \bullet NINT\left(\frac{f}{g}\right) \right|, \tag{4.3}$$

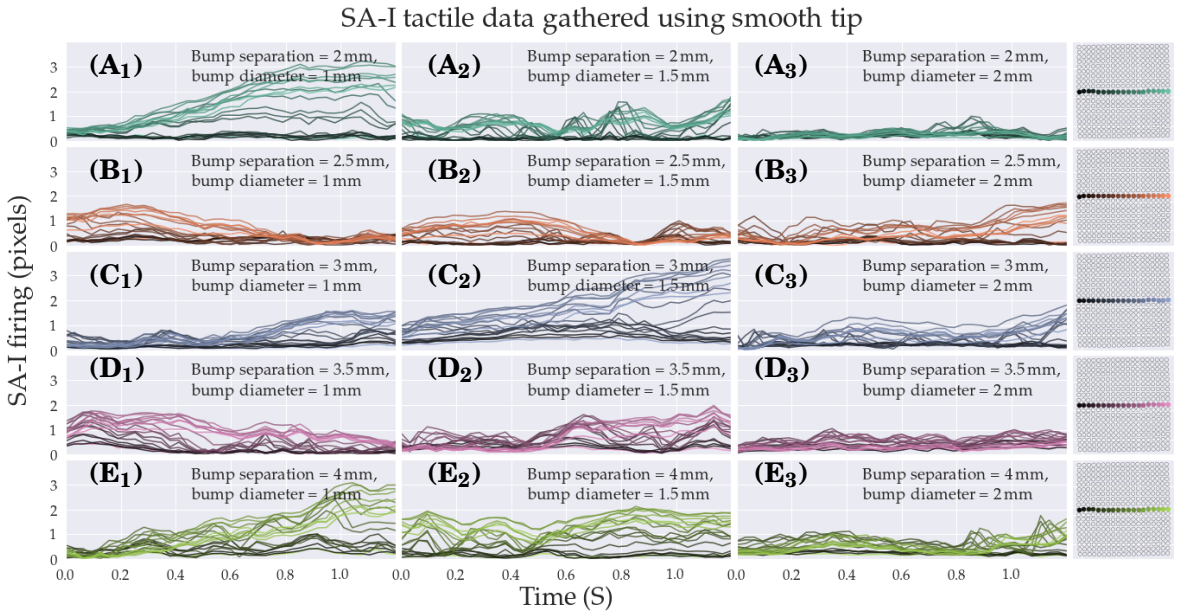

Figure 4.21: 1.2 s samples of artificial SA-I firing collected using a dynamic touch with the smooth tip at 50 mms^{-1} on stimuli described in 4.7.1.

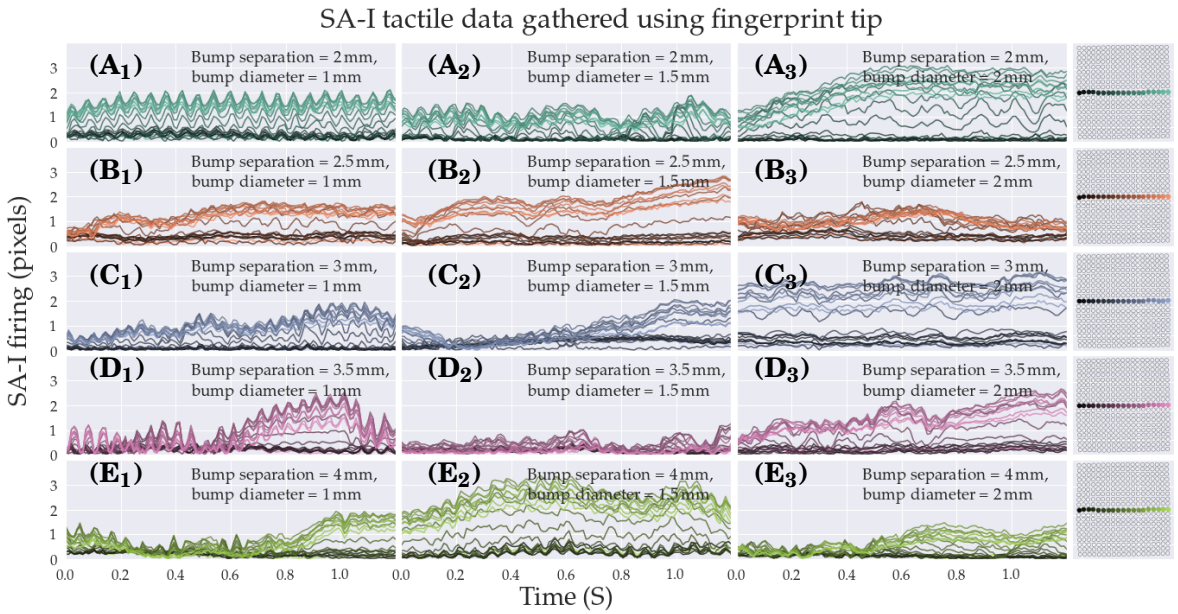

Figure 4.22: 1.2 s samples of artificial SA-I firing collected using a dynamic touch with the fingerprint tip at 50 mms^{-1} on stimuli described in 4.7.1.

where *NINT* is the the nearest integer function. So, for example, with the smooth tip, we have a low frame-rate of ~30 Hz, if the true signal is 20 Hz it will be reconstructed at 10 Hz in the measured signal. This is major source of error for the smooth tip, however, aliasing will also take effect in the fingerprint tip although the threshold (Nyquist) frequency is higher. Despite this we

attempt to make some tentative comparisons between responses for the two tip types: Aside from the aforementioned DC-bias, SA-I response for the fingerprint tip showed considerable periodicity, with periods as low as ~0.05 s, whereas, we find that the smooth tip exhibited no such periodicity. This may be indicative of the effect of artificial papillary and epidermal ridges seen by Scheibert et al. (2009) [59] described in 2.1.1. If we interpret our results in this way, the smooth tip can be considered a low-pass filter which attenuates high-frequencies whilst the fingerprint tip allows these higher frequencies to be transmitted.

Scheibert et al. observed that their artificial papillary ridges created a bandpass filter, amplifying frequencies around the spatial frequency $1/\lambda_{\text{fp}}$, where λ_{fp} was the spatial period of their artificial fingerprint (220 μm). Applying the findings of Scheibert et al. to our experiments, the theoretical amplified frequency, $f_A = v/\lambda_{\text{fp}}$, where v = 50 mms^{-1} is the scanning velocity and λ_{fp} = 3 mm is the inter-nodule distance in the scanning direction, is ~17 Hz. The typical frequencies observed in Fig. 4.22 are ~20 Hz so it is be reasonable to judge that the amplification observed in our results are a consequence of the effect described by Scheibert et al. However, we must also consider that the reduced sampling rate attained with the smooth tip is likely to cause aliasing of any frequencies above f_{Nyquist} =~15 Hz. Therefore, at this stage, we cannot say with confidence that the fingerprint bandpass effect is being observed here.

According the Scheibert et al., at natural exploratory scanning speed (10-15 cms^{-1}) papillary ridges on the human fingertip will amplify frequencies of ~300 Hz. Therefore, it is unlikely natural SA-I or FA afferents will respond to these amplified frequencies, however, ~300 Hz is in the range of preferential sensitivity of PC afferents. Thus, an important consequence of the physical dimensions used in our experiment is that the consequential amplification may be observed in the SA-I signal which has a theoretical maximum frequency of f_{Nyquist} =~45 Hz for the fingerprint tip.

4.7.2.2 FA

Figs. 4.23 and 4.24 show modelled FA firing, for smooth and fingerprint tips respectively, in the same 1.2 s sample as Figs. 4.21 and 4.22, when the TacTip has undergone stimulation as described in 4.7.1.

As with artificial SA-I firing, a visual inspection of the data provides no obvious systematic relationship between signals and the objective stimulus quantities (bump separation and bump diameter). Unlike SA-I afferents, artificial FA afferents did not exhibit any DC offset. This is by virtue of the adaptation rate: sustained or slow changes in shear forces resulting from overall drum diameter do not appreciably affect marker velocities.

Unfortunately, as with artificial SA-I, we observe the consequence of a reduced frame-rate with the smooth tip. Nevertheless, we attempt to make some comparison between the two tip types: For the smooth tip (Fig. 4.23), rather than any DC-component, we observe what appears to be random noise, with no coherence across afferents. For the fingerprint tip (Fig. 4.24), in some

cases ($\mathbf{A_1}$, $\mathbf{C_1}$ and $\mathbf{D_1}$) we observe periodic peaks with the same period as the high frequency components in Figs. 4.22A_1, 4.22C_1 and 4.22D_1. Any signal produced by other stimuli was not strong enough to be observed above the what appears to be noise, with an amplitude of ~0.1 pixel/s. It is possible that the difference in output between smooth and fingerprint tip variants was a consequence of the filtering effect described by Scheibert et al. [59], however, as with SA-I, we must also consider the possibility of sampling rate discrepancy in causing the observed differences.

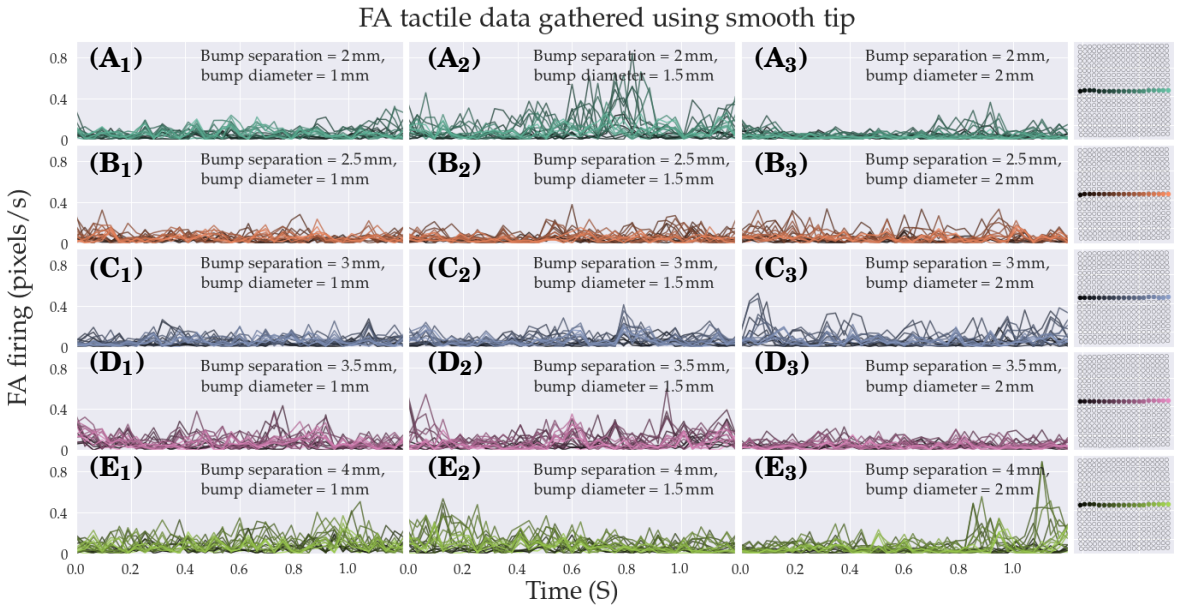

Figure 4.23: 1.2 s samples of artificial FA firing collected using a dynamic touch with the **smooth tip** at 50 mms^{-1} on stimuli described in 4.7.1..

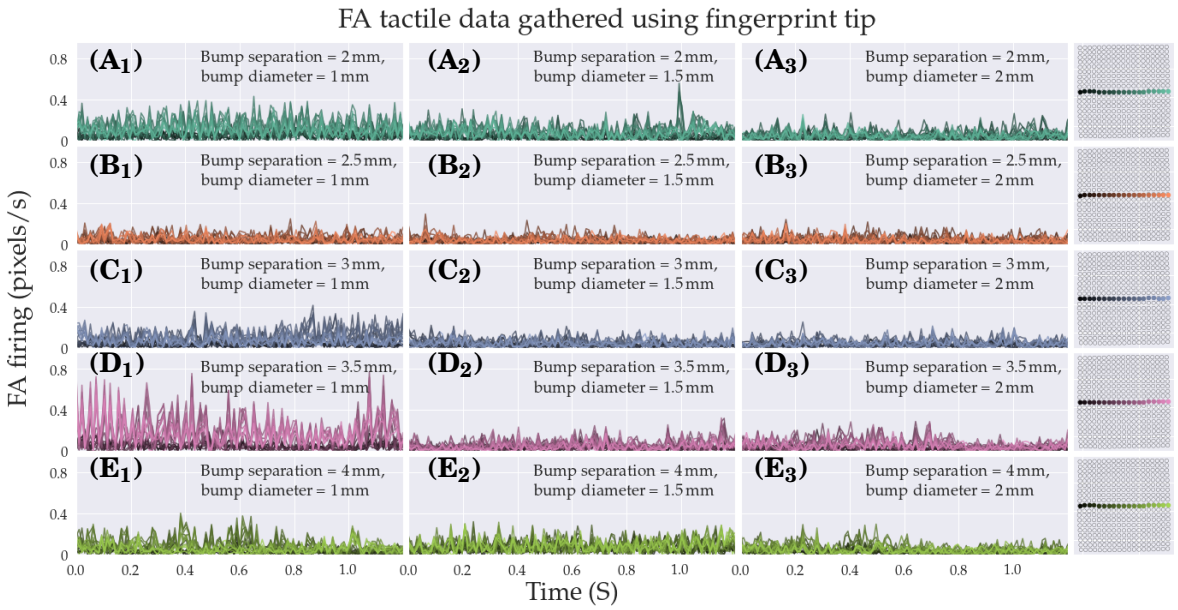

Figure 4.24: 1.2 s samples of artificial FA firing collected using a dynamic touch with the **fingerprint tip** at 50 mms^{-1} on stimuli described in 4.7.1.

4.7.2.3 Vibrational Channel

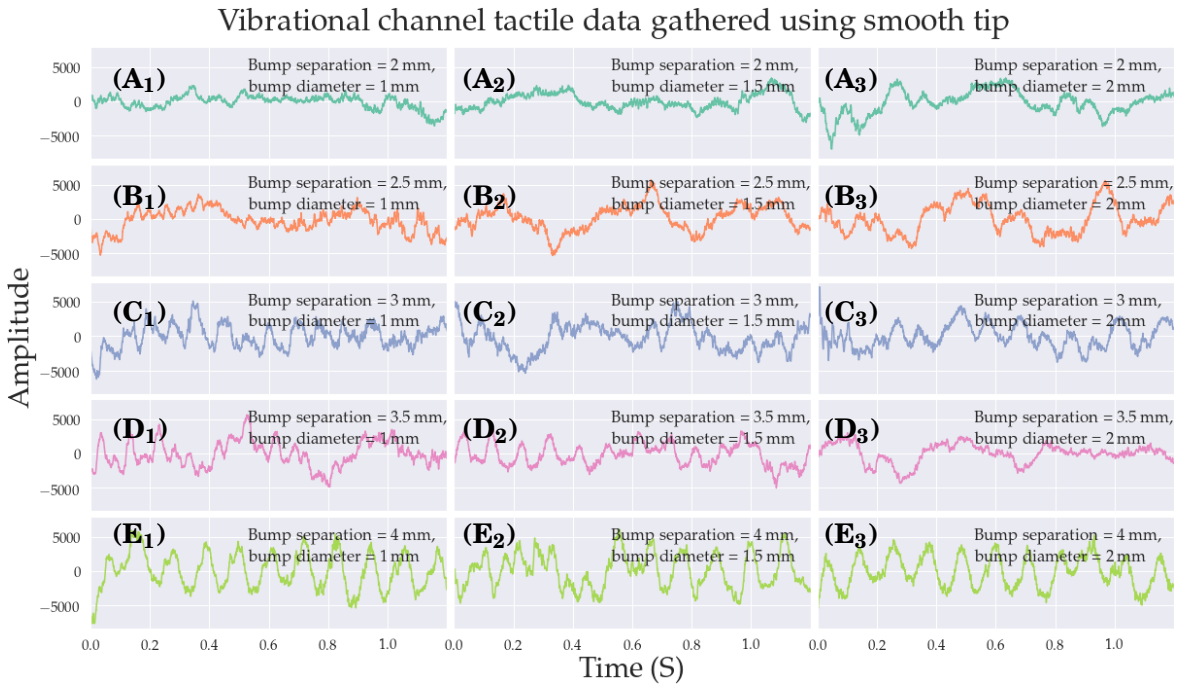

Figure 4.25: 1.2 s samples of vibrational data collected using a dynamic touch with the smooth tip at 50 mms^{-1} on stimuli described in 4.7.1.

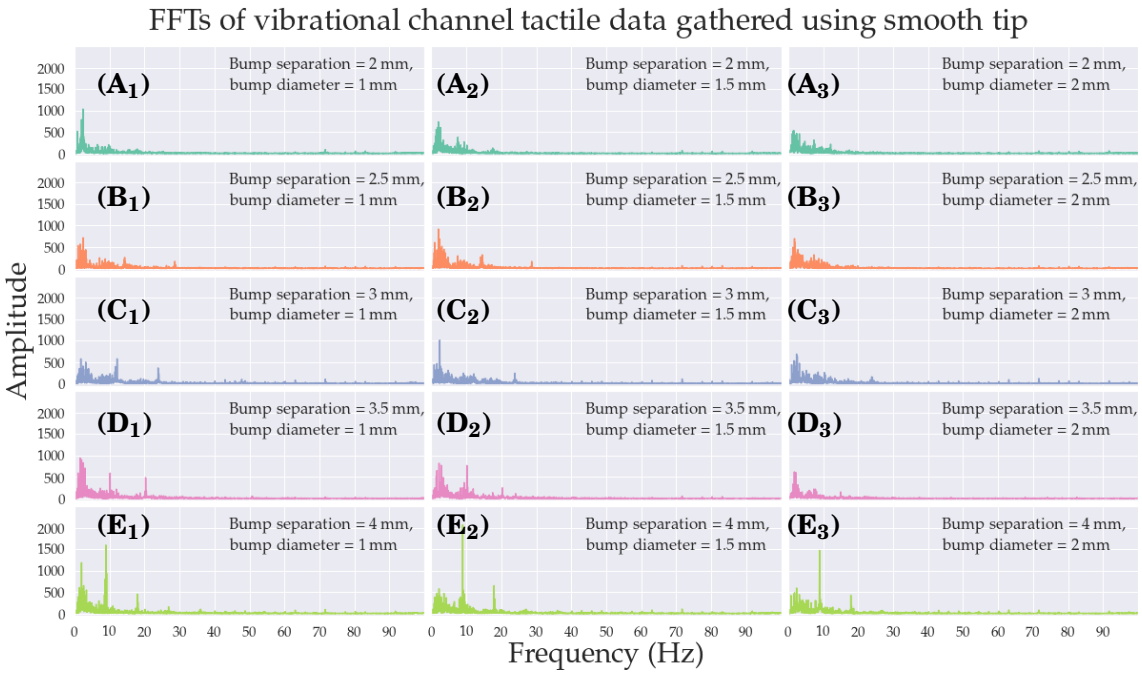

Figure 4.26: Fast Fourier transforms for tactile vibrational data collected using a dynamic touch with the smooth tip at 50 mms^{-1} on stimuli described in 4.7.1.

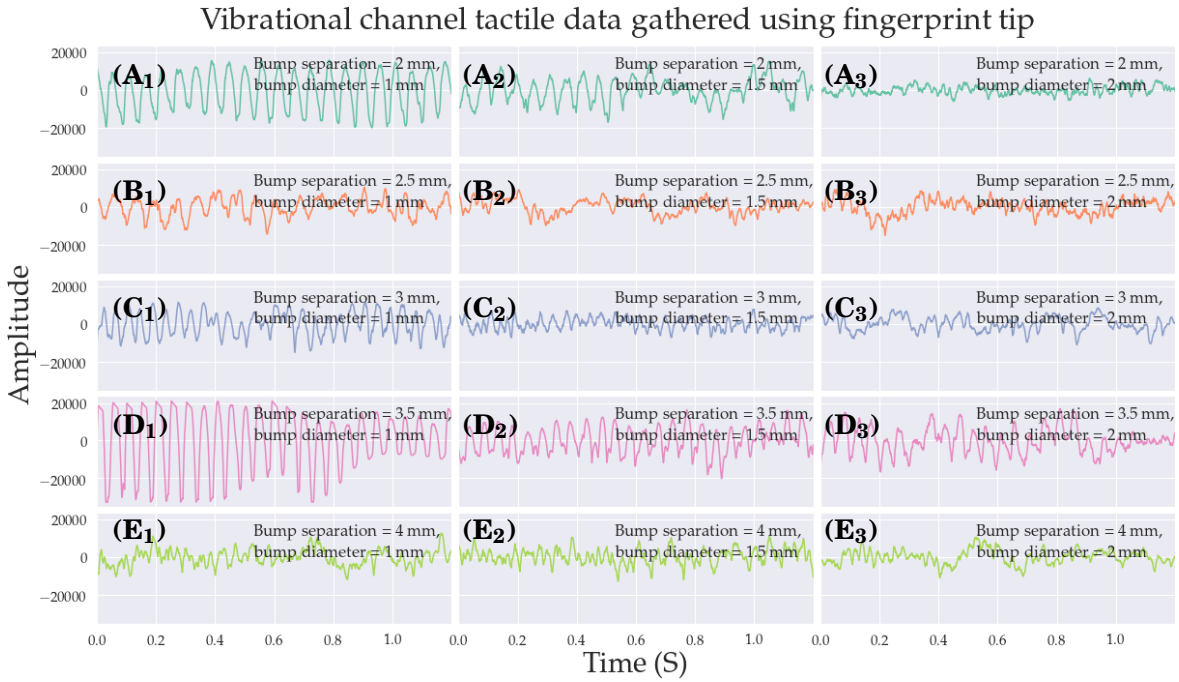

Figure 4.27: 1.2 s samples of vibrational data collected using a dynamic touch with the fingerprint tip at 50 mms^{-1} on stimuli described in 4.7.1.

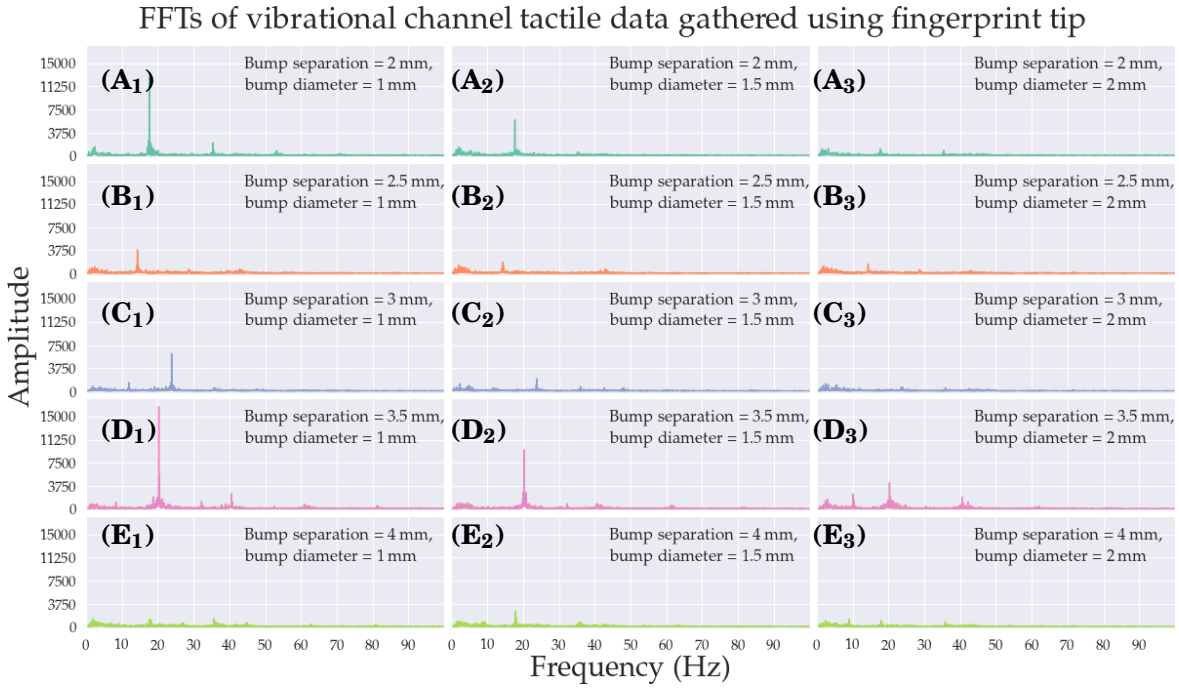

Figure 4.28: Fast Fourier Transforms for tactile vibrational data collected using a dynamic touch with the fingerprint tip at 50 mms^{-1} on stimuli described in 4.7.1.

Figs. 4.25 and 4.27 show vibrational data collected with the TacTip, for smooth and fingerprint tips respectively, in the same 1.2 s sample as Figs. 4.21 and 4.22. Figs. 4.26 and 4.28 show fast Fourier transforms (FFTs) performed on the whole 30 s period of tactile vibration data for smooth and fingerprint tips respectively. It is worth noting that, with the vibrational channel, unlike SA-I and FA, the sample rate is the same for both tips. For both time and frequency domain data, the results are presented in arbitrary units because the relationship between the voltage level recorded by the PC soundcard and the number reported by python sound-device library is not known, furthermore, the relationship between a unit of audio amplitude (e.g., decibels) and the voltage produced by the elecret microphone is also unknown. For the purpose of this work, however, we are concerned with the relative aspects of vibrotactile data; i.e., how the data collected on different stimuli compares.

Natural frequencies (fundamentals), $f_{\text{natural}} = v/w$, where $v = 50\,\text{mms}^{-1}$ is the stimulus velocity and $w = x\sqrt{2}/2$ is the inter bump spacing in the direction of travel (see Fig. 4.19), are ~36, 28, 24, 20, and 18 Hz for stimuli A*, B*, C*, D* and E* respectively. It is possible that, as well as this natural frequency, there is a second natural frequency, $f_{2,\text{natural}} = 0.5 f_{\text{natural}} = v/2w$, owing to the bump spacing associated with bumps which fall directly in-line with one-another in the direction of travel. Let us first consider the FFTs produced with the smooth tip (Fig. 4.26). For each stimuli, apart from $\mathbf{A_1}$, $\mathbf{A_2}$ and $\mathbf{A_3}$, we see the natural frequency as a small peak. We do

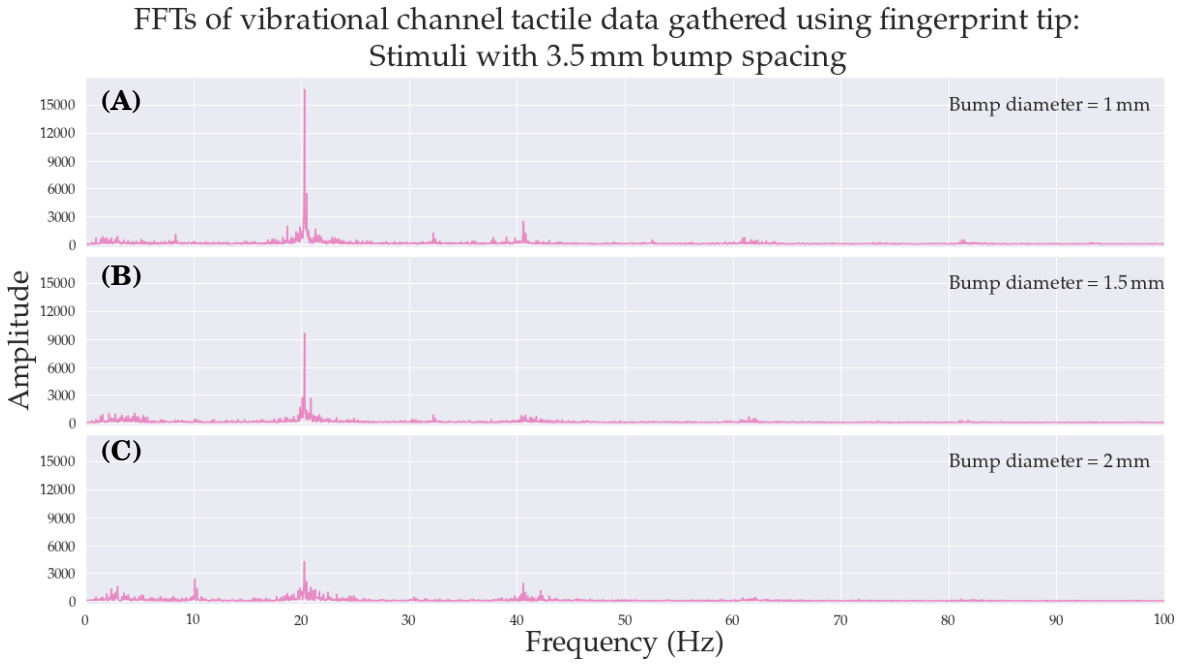

Figure 4.29: FFTs for tactile vibrational data collected using a dynamic touch at $50\,\text{mms}^{-1}$ on stimuli with a 3.5 mm bump separation and bump diameters of 1 ($\mathbf{D_1}$), 1.5 ($\mathbf{D_1}$) and 2 mm ($\mathbf{D_1}$): Figs. **A**, **B** and **C** respectively.

also observe peaks at $f_{2,\text{natural}}$ in all stimuli apart from $\mathbf{A_1}$, $\mathbf{A_2}$ and $\mathbf{A_3}$. In-fact, the highest peak is observed for stimuli $\mathbf{E_1}$, $\mathbf{E_2}$ and $\mathbf{E_3}$ at 9 Hz. It is possible that the 2$^{\text{nd}}$ harmonic of $f_{2,\text{natural}}$ consistently contributes to the peak at f_{natural}

Now considering the fingerprint tip (Fig. 4.28), in contrast to the smooth variant, natural frequencies, f_{natural} and $f_{2,\text{natural}}$ are observed for all four bump separations and the amplitudes are generally much greater. Additionally, in some cases, harmonics can be observed; e.g., at $3f_{2,\text{natural}}$ for stimulus $\mathbf{A_1}$ and $2f_{\text{natural}}$ for stimuli $\mathbf{D_1}$, $\mathbf{D_2}$ and $\mathbf{D_3}$. For each bump separation, the natural frequency or harmonic which is closest to ~20 Hz was the most powerful. These results appear to confirm that the smooth tip acted as a low-pass filter, attenuating frequencies higher than ~10 Hz and that the fingerprint tip acted as a broad bandpass filter about ~20 Hz. As with the SA-I and FA response, these filtering properties agree with the theory proposed by Scheibert et al. [59], that a fingerprint operates as a bandpass filter about spatial frequencies λ_{fp}, where λ_{fp} is the distance between fingerprint ridges or nodules.

Now considering only the fingerprint tip, we find that overall spectral power was inversely proportional to bump diameter; i.e., for a given bump separation, the amplitude of each natural harmonic decreased as bump diameter increased. To exemplify this, we show three FFTs collected on stimuli with a bump separation of 3.5 mm at Fig. 4.29. Notice how the most prominent harmonic, f_{natural}, present at all three bump diameters, has a reduced relative amplitude as bump diameter increases. This indicates that whilst frequency may provide a cue for bump separation, intensity may provide a cue for bump diameter. Of course, intensity co-varied with bump separation because of the relative position of the natural harmonics to the λ_{fp} band of amplification, thus intensity of vibrations alone is not a universal cue for bump diameter. Methods that consider both the harmonic frequencies and their relative amplitudes, however, are likely to be successful in providing perceptual judgements about textures that vary over both of these dimensions: it is believed that harmonic structure of vibrotactile data provides a cue for texture discrimination in nature [131, 133], perhaps, therefore the overall shape of these FFTs can be decoded for artificial perception of the stimuli used in this experiment.

4.7.3 Discussion

For SA-I, FA and vibrational channels, the addition of an artificial fingerprint appeared to make an appreciable difference to the signals. It appeared to amplify salient information in both the case of artificial SA-I and FA afferents, although it was difficult to isolate the amplification effects from an unfortunate side-effect of a reduced frame-rate when using the smooth tip. Upon examining the vibrational channel data, it appears that the fingerprint serves as a bandpass filter around the spatial frequency $1/\lambda_{\text{fp}}$.

On the stimulus scales used here, the results suggest that, with the fingerprint tip, the vibrational channel detects sufficient information to make perceptual judgements regarding the textures: information relating to bump separation is contained in the frequencies of natural

harmonics and information relating to bump diameter is contained in the intensity of the signal. It seems plausible that more complex, natural textures might contain superimposed natural frequencies and associated harmonics. Our hypothesis, therefore, is that *harmonic structure* of induced vibrations could provide a cue for texture discrimination. In fact, this agrees with a compelling theory of human texture perception which aims to explain the observed invariance to scanning speed [131]. The theory is based on an idea that touch leverages mechanisms for processing spike timing, thought to enable timbre invariance in hearing [133], which is based on the fact that harmonic structure is consistent across scanning speeds.

The peak frequencies recorded by the vibrational channel with the fingerprint tip were well within the Nyquist frequency for artificial SA-I and FA channels, $f_{Nyquist}$. In fact, by comparing the vibrational channel data in the time domain, with that of the SA-I response, we notice that the general form of each signal is similar in terms of amplitude and frequency. This is perhaps unsurprising since the vibrational channel is measuring acoustic vibrations in the gel which have been produced by the skin dynamics, which marker positions directly record. Therefore, we believe that, on the stimulus scales used here, SA-I and FA channels may also be useful in perceiving texture.

In theory, as bump separation decreases, the natural frequency will increase until it exceeds the Nyquist frequency of SA-I and FA channels, at which point the vibrational channel may be required. However, the transduction of higher frequencies is limited by the scale of the fingerprint: the theoretical amplification band provided by our artificial fingerprint at these speeds is ~20 Hz, thus, if we were to reduce bump separation of the stimuli, we expect that the fingerprint scale should also be reduced in order to amplify these higher frequency harmonics. Indeed, in the bandpass theory of natural fingerprints [59], papillary ridge spacing of 220 μm produces amplification with a bandpass filter located at ~300 Hz (preferential sensitivity of natural PC channel) at normal exploratory speeds. Unfortunately the scales of our artificial fingerprint and stimuli are somewhat limited by the tolerances of the 3D-printing technologies available.

4.8 Concluding Remarks

In this chapter, we presented three main hardware modifications to the TacTip: *(1)* addition of a *vibrational channel* via an electret microphone embeded in the gel; *(2)* a novel marker arrangement intended to more closely mimic the innervation density of SA-I and FA afferents to the human fingertip; *(3)* addition of an artificial fingerprint, intended here to amplify vibrations for the purpose of texture perception.

We present two novel feature extraction methods, derived from TacTip marker positions, which are proposed to model natural SA-I and FA afferents. We also suggest that the vibrational channel mimics the innervation density and response characteristics of the natural PC channel

and therefore may be leveraged for perception of tactile dimensions for which the PC channel is believed to be responsible; e.g., fine texture.

The response of artificial SA-I and FA afferents, when the TacTip was stimulated with transient and sustained normal pressure, was indicative of the firing rates of their natural analogues. In particular, artificial SA-I afferents continued to respond to sustained pressure whilst artificial FA afferents only responded to changes in pressure. This is coincidental of the adaptation rates of natural afferents.

The response of artificial SA-I and FA afferents to aperiodic gratings exhibited similarities with natural counterparts in monkeys. Notably, both artificial and natural SA-I afferents were preferentially sensitive to edges over ridges. Also, SA-I and FA artificial afferent response showed similar modulation with gap width to their respective natural afferents.

The detail shown in natural SA-I afferent spatial response profiles (SRPs) [112] suggest a similar limit on the spatial resolution to that observed in psychophysical studies [43] (diminishing accuracy $<\sim 2$ mm), leading the authors to conclude that information about spatial properties of the stimulus are mediated via modulation of firing rates across populations of SA-I afferents. Artificial SA-I SRPs were significantly less detailed suggesting that the spatial resolution of this channel is considerably less than its natural analogue: when gaps were closer than ~ 4 mm artificial afferents consistently responded to individual ridges and edges as though they were single flat surfaces.

The authors of [112] were able to rule out FA afferents as a candidate for mediating information about stimulus shape via a spatial code, when pressing stimulation is employed, owing to a relative lack of detail in their SRPs. Although, the authors did not investigate sliding stimulation. The difference between artificial FA and SA-I afferents was less pronounced and therefore we predict that their capacity for mediating spatial codes of stimulus shape may be comparable.

Despite the clear similarities in response profiles between natural and artificial afferents, there are constraints to the likeness of the presented model to its natural counterpart. As stated in Section 4.6.4, it is likely that natural afferents are sensitive to normal forces (compressive strain of skin) [282], a property to which marker positions and speed, as viewed by the camera, do not respond. Furthermore, mechanoreceptors are known to respond to local deformation rather than global skin displacement whereas, here the measures taken to model SA-I and FA afferent response are in a fixed global frame of reference; i.e., the camera. One possible solution to this shortcoming is to use a measure that captures relative marker displacements; e.g., the Voronoi approach to feature extraction developed for use with the TacTip [283]. We opted against this approach since, with it, the established analogy between the morphology of Merkel cells complexes in relation to papillary ridges and markers in relation to pins [12] is lost.

Finally, we made a qualitative assessment of the capacity for the three artificial tactile channels to mediate information about texture via dynamic stimulation. Importantly, by comparison with a 'smooth tip', we found that addition of an artificial fingerprint produced a more pronounced

periodic response in all three tactile channels. We hypothesised that this was the same effect as observed by Scheibert et al. (2009) [59] that the fingerprint tip creates a bandpass filter around the spatial frequency $1/\lambda_{\text{fp}}$, where λ_{fp} is the spatial period of the fingerprint, whilst the smooth tip attenuates lower frequencies.

The bandpass effect was most obvious for vibrational channel data, where we saw natural frequencies of each texture and associated harmonics related to the spatial period of bumps. The amplitudes of peak frequencies in the vibrational channel was inversely proportional to bump diameter. These results led us to hypothesise that harmonic structure of induced vibrations may provide a viable cue for the set of textures used here.

Neither artificial SA-I or FA channels exhibited any obvious systematic relationship between signal and objective stimulus quantities (bump separation and bump diameter). However, time domain data from the vibrational channel was quite similar to that of SA-I response in the same period. Furthermore, the peak frequencies recorded by the vibrational channel were below the Nyquist frequency for the SA-I and FA channels which suggests that SA-I and FA data may contain information that could be useful for the perception of texture on the scales used here.

This study is by no means the first to draw inspiration from the tactile channels in human touch. For example, TacTip markers have been proposed as analogous to Merkel cell complexes (SA-I) [12]; Romano et al. (2011) [262] loosely mimicked SA-I, FA and PC channels from capacitive array outputs of 5x3 taxels and accelerometers; and the BioTac [256] has been suggested to possess artificial channels analogous to SA-I and PC. However, where the work presented in this chapter builds upon previous examples of bio-inspired tactile sensors and contributes significantly to the field of biomimetic touch, is in the attempt to directly model an abstract neural code (rate code) of SA-I and FA afferents. Furthermore, these artificial neural codes were directly compared to natural firing rates of SA-I and FA afferents in monkeys under the same stimulation conditions, which, as stated, produced remarkably similar results (Section 4.6.3). To our knowledge, this is the first study to directly compare artificial neural codes, derived from a digital tactile sensor, to natural neural response. The methodology presented for performing these experiments and, the analysis of results in terms of artificial skin dynamics and the properties of which individual artificial afferents encode, is likely to inform future work towards even more biologically plausible artificial tactile sensors.

Recently a branch of tactile sensors described as being 'neuromorphic' have been subject to increased research and development. In these sensors, fundamental spike-based signals of natural tactile channels are directly modelled; e.g., using biological models of SA-I [263, 264] and FA [264, 265] neurons or using neuromorphic cameras in adapting optical tactile sensors to produce spiking output [284]. Importantly, the spike-based output of these devices is continuous in time; i.e., there is no associated sample rate. This theoretically enables stimulus information to be contained within any type of neural code which has been hypothesised in nature. Most notably, these devices permit temporal encoding based on precise inter-spike timing [73]. In

contrast, in the work presented here, we have stated that we are modelling an abstract rate code; i.e., each camera frame models an encoding window [73], within which the number of SA-I and FA spikes is modelled by marker displacement and speed respectively. It is out of necessity that we model this abstract code since the type of data (continuous values (marker dynamics) and discrete samples (frames)) does not enable inter-spike timing temporal codes. This is a potential limitation of our approach, that can only be solved by using neuromorphic technologies, both in terms of the capacity to capture relevant stimulus information and also in the degree to which we can accurately model natural systems, since many believe neural code based on inter-spike timing to be significant in nature [266]. However, in the subsequent chapter we describe how a temporal code can built on top of the rate code model presented in this chapter, which is based rate fluctuations; i.e., temporal variations in firing rate between encoding windows, an encoding mechanism which is also classed as a type of temporal code in neuroscience [74]. Furthermore, by assuming this abstract model of neural code, our approach is much simpler than neuromorphic methods and can be applied to conventional machnine learning models for decoding spatio-temporal signals where neurmorphic sensors require decoders with time dependencies such as spiking neural networks.

It is worth mentioning here, that a number of other sensors offer alternatives to the engineered vibratoinal channel, in particular for use with texture sensing [149, 183, 195, 216, 217, 219, 222, 227–230]. These have been extensively described in the literature review, Section 2.2.2.6. It is difficult to make any meaningful comparison to the alternatives at this stage, since we are yet to apply the vibrational channel to a texture recognition task, so the bulk of our discussion in relation to where this aspect of the present study sits within the field will be made in the following chapter. However, in the present chapter it was suggested that harmonic structure of the vibrational channel data might be a viable code for texture perception so it is interesting to comment on how this relates to other similar studies within the field: By far the most common approach to feature extraction for artificial texture perception is to use statistical features extracted from frequency domain data [183, 216, 217, 219, 229]. More recently, the emergence of deep-learning has enabled the use of higher-dimensional feature sets where raw data in the time-domain is leveraged to predict texture [228, 229] achieving high accuracy and good generalisation on complex natural surfaces. The statistical features used in aforementioned studies are of course implicit within the harmonic structure of texture induced vibrations and given the success of deep-learning methods employed within the time-domain, we predict that these methods may be able to extract relevant features from harmonic structure in-place of hand crafted statistical features.

In the following chapter, we use these artificial tactile channels in a texture discrimination task under two stimulation regimes: static and dynamic touch. SRPs of artificial SA-I and FA afferents (Section 4.6.3) suggest that these artificial channels provide sufficient resolution to spatially resolve coarse textures with static touch. The level of detail in artificial SA-I SRPs was less than that of monkey SA-I afferents [112] suggesting that the limit on spatial resolution may

be higher for our system than that of humans. FFTs of vibrational channel data suggest that, in agreement with a hypothesis about human tactile texture perception [131], harmonic structure of induced vibrations is a viable cue for texture discrimination.

CHAPTER 5

ARTIFICIAL AFFERENTS APPLIED TO TEXTURE PERCEPTION

In this chapter we assess the ability of the artificial tactile channels, described in Chapter 4, to be used for texture classification. We use two modes of stimulation: *static* and *dynamic* and examine a number of encoding schemes inspired by theories of neural encoding in natural afferents for human texture perception.

5.1 Background

In the previous chapter, we presented three artificial tactile channels of the TacTip, SA-I, FA and a vibrational channel. We suggested that artificial SA-I and FA output are viable models for their natural analogues based on comparisons to physiological data (Section 4.6.3) and therefore have the ability to spatially encode tactile stimuli, for which natural SA-I and FA afferents are believed to be responsible [43]. We constructed hypotheses of their capacity to mediate spatial cues on different scales (artificial SA-I and FA) and considered the capacity for all three channels to be leveraged for dynamic texture discrimination through qualitative analysis of their data. Here we assess the performance of artificial SA-I and FA channels for texture discrimination when statically stimulated and all three tactile channels when dynamically stimulated.

Our literature review of human texture perception (Section 2.1.3.6) alluded to the complexity of this tactile dimension. In particular, it appears that humans naturally employ two stimulation regimes depending on the spatial scale of the texture: static touch for coarse textures and dynamic touch for fine [57]. This phenomenon was theorised by Katz (1925) [125] which he dubbed 'duplex theory'. There is a theoretical limit on the ability to perceive textures with a static touch based on the idea that spatial cues are mediated via a spatial code of SA-I and possibly FA afferents [110, 129]. The innervation density of these type-I afferents in the human fingertip presents a

limit on the ability of these fibres to mediate information via a spatial code.

Phillips and Johnson (1981) [112] looked at SA-I and FA afferent discharge in monkeys when indented with aperiodic gratings of varying ridge and gap widths. The purpose was to investigate the degree to which spatial resolutions observed in a preceding psychophysical paper [43] can be explained by modulation in afferent discharge rates with varying spatial scales of the stimulus. Spatial response profiles (SRPs) suggested that, when the space between ridges was reduced below a certain distance (~1 mm), neither SA-I or FA afferents exhibited enough detail to resolve the stimulus shape. In the preceding chapter, we performed an identical study with artificial SA-I and FA afferents (Section 4.6) and found that ~4 mm was the analogous limit on spatial resolution of artificial SA-I and FA afferents. We therefore hypothesise that, where spacing of textural elements are below ~4 mm, the performance of static models of texture perception based on the spatial codes of artificial SA-I or FA firing will be drastically reduced.

Where textures are too fine to be spatially resolved with type-I afferents via static touch, humans naturally employ dynamic stimulation, involving 'sliding' a finger relative to texture. It is believed that induced vibrations provide relevant cues for texture perception under these conditions [57, 126] and the primary channel for transducing this information is PC [126–128] although some evidence suggests that also FA afferents may mediate frequency based cues for texture perception [129]. An interesting phenomenon related to dynamic texture perception is speed-invariance [126], where the effect of sliding speed is not significant on perceived roughness, as might be predicted with the use of simple frequency based cues. Boundy-singer et al. (2017) [131] suggest a possible mechanism for speed-invariance: cortical computations are capable of extracting *harmonic structure*, which is preserved across speeds, from neuron firing, the frequency composition of which reflects that of the oscillations elicited in the skin during texture scanning [88], implying a temporal neural code. We therefore hypothesise that harmonic structure of data collected via the engineered vibrational channel may be a viable cue for texture discrimination, particularly for fine textures.

We assess the ability for artificial SA-I and FA channels to mediate relevant textural cues via a spatial code using convolution neural networks (CNNs) trained on static tactile images of each afferent. We propose that the level of performance of these CNNs in a texture discrimination task will indicate the degree to which relevant spatial information is encoded within tactile images of artificial SA-I and FA afferents. We also assess the ability for artificial SA-I and FA channels to mediate relevant textural cues via a spatio-temporal code, suggesting that modulation in artificial firing rate across a 10-frame encoding window, collected with dynamic sliding, may contain relevant information. In this case, LSTMs are used to extract possible temporal features. Finally, we assess the ability for the novel vibrational channel to mediate vibrational cues that correlate with texture. In accordance with the theory that *harmonic structure* provides a speed-invariant cue for texture discrimination [131], we use frequency spectra as samples for texture prediction and use 1D-CNNs to extract structural features.

5.2 Data Collection

5.2.1 Robotic System and Software Structure

Table 5.1: Complete set of stimuli used. Refer to Fig. 4.19 for dimensinos $\boldsymbol{x}$ and $\boldsymbol{d}$.

Stimulus		Dimensions (mm)
a		$x = 0$, $d = 0$
b		$x = 0.5$, $d = 0.5\sqrt{2/\pi}$
c		$x = 1$, $d = \sqrt{2/\pi}$
d		$x = 1.5$, $d = 1.5\sqrt{2/\pi}$
e		$x = 2$, $d = 2\sqrt{2/\pi}$
f		$x = 2.5$, $d = 2.5\sqrt{2/\pi}$
g		$x = 3$, $d = 3\sqrt{2/\pi}$
h		$x = 3.5$, $d = 3.5\sqrt{2/\pi}$
i		$x = 4$, $d = 4\sqrt{2/\pi}$
j		$x = 4.5$, $d = 4.5\sqrt{2/\pi}$
k		$x = 5$, $d = 5\sqrt{2/\pi}$
l		$x = 5.5$, $d = 5.5\sqrt{2/\pi}$
m		$x = 6$, $d = 6\sqrt{2/\pi}$

We employ the same robotic system as in Chapter 4. For details of the tactile sensor please refer to Sections 4.2.1 and 4.2.2. For details of the robot arm and software structure please refer to Section 4.3.

For the work in this chapter we use the fingerprint tip, described in Section 4.2.2.4. This is because it was found to amplify transient signal (Section 4.7) and to have little affect on spatial resolution as elucidated by the grating experiment in Section 4.6.

5.2.1.1 Stimuli

Throughout this chapter we use a set of $N_{\text{tex}} = 13$ (***a***-***m***) textured stimuli. As in Section 4.7, stimuli consist of tetragonal arrays of raised bumps. The dimensions, x and d (defined in Fig. 4.19) vary according Table 5.1. x and d were chosen to provide a linear variation in texture period whilst maintaining the area of raised bumps across stimuli. The purpose of keeping the area of raised bumps consistent was to prohibit the availability of intensive cues.

We employ two distinct modes of stimulating the TacTip: *i)* pressing, where the stimuli are fabricated on square plates (Fig. 5.1), 3D-printed in rigid plastic (VeroWhite, Stratasys), and *ii)* sliding, where the textures are 3D-printed in the same way on a set of 13 drums (diameter = 80 mm) which are used to dynamically stimulate the TacTip as in Section 4.7.1.

5.2.2 Pressing

The robot makes successive taps onto the flat textured stimuli indenting the skin by ~1 mm (see Fig. 5.1). Each tap takes ~1 s; the robot is stationary at the bottom of a tap for 0.2 s. TacTip video data is simultaneously recorded for the entire downward phase of each press.

On each discrete tap, the robot's x, y and z positions are randomly varied according to uniform distributions in the ranges ±2, ±2 and ±0.7 mm respectively. Roll, pitch and yaw angles are also varied randomly on each tap according to uniform distributions in the ranges ±3.6, ±3.6 and ±90° respectively. The purpose of varying these co-ordinates is to avoid providing positional or rotational cues which may otherwise be implicitly available through the experimental set-up, for example, inconsistency in 3D-printing or how each stimulus is fastened to the test bed. In this way, spatial cues relating to textural qualities of the stimulus are isolated.

750 taps per stimulus class are performed, yielding $N_{\text{samples}} = 750$ samples per class and $N_{\text{tex}}N_{\text{samples}} = 9750$ samples in total.

5.2.3 Sliding

The TacTip, mounted as an end-effector to the ABB arm, is held stationary against textured drums (see Fig. 4.20). A custom built rig rotates the drums in order to dynamically stimulate the TacTip (see Section 4.7.1.2 for details of the rig).

The drum is indented ~1 mm into the TacTip's skin by the ABB robot arm. For each drum the TacTip is stimulated at a set of $N_{\text{speeds}} = 10$ linear speeds, linearly spaced from 10 to 100 mms^{-1}, yielding a total of $N_{\text{tex}}N_{\text{speeds}} = 130$ stimulus conditions. The TacTip is stimulated for 65 s per stimulus condition whilst simultaneously recording TacTip video and audio data. The initial and final 2.5 s of each recording are thrown away leaving only constant speed stimulation.

5.3 Overview of Competing Models

This brief section serves as a reference for the approaches to robotic texture perception subsequently described and tested.

Table 5.2: Table of reference for the developed models of robotic tactile texture perception.

Model	Tactile channel	Data	Train/test	Encoding
SA1-STC-*press*	SA-I	pressing	-	spatial
FA-STC-*press*	FA	pressing	-	spatial
SA1-STC-*slide*	SA-I	sliding	*all speed testing*	spatial
FA-STC-*slide*	FA	sliding	*all speed testing*	spatial
SA1-STC-*slide*-$\boldsymbol{v}$	SA-I	sliding	*leave-one-speed-out*	spatial
FA-STC-*slide*-$\boldsymbol{v}$	FA	sliding	*leave-one-speed-out*	spatial
SA1-DTC	SA-I	sliding	*all speed testing*	spatio-temporal
FA-DTC	FA	sliding	*all speed testing*	spatio-temporal
vibro-DTC	vibrational	sliding	*all speed testing*	temporal
SA1-DTC-$\boldsymbol{v}$	SA-I	sliding	*leave-one-speed-out*	spatio-temporal
FA-DTC-$\boldsymbol{v}$	FA	sliding	*leave-one-speed-out*	spatio-temporal
vibro-DTC-$\boldsymbol{v}$	vibrational	sliding	*leave-one-speed-out*	temporal

Twelve classes of model are designed and tested. Six of these model types are characterised as static and the remaining six are characterised as dynamic. The fundamental difference being that only dynamic methods permit temporal encodings of texture.

At this stage, it is important to make the distinction between the terms '**sliding**' and '**dynamic**'. In this work, **sliding** refers to the method used for data collection (Section 5.2.3), whereas **dynamic** refers to a class of model. Importantly, there are static as well as dynamic models which use data collected via sliding; i.e., in terms of the methods developed, the use of sliding data and static models are not mutually exclusive.

For those models which use sliding data, there are two regimes for training and testing: *i) All speed testing* and *ii) Leave-one-speed-out cross validation*. These approaches are subsequently explained in more detail but, essentially, with *all speed testing*, 1 model is constructed which has been trained and tested with examples of data collected at each of the 10 speeds and with *leave-one-speed-out cross validation*, 10 models are constructed, each one trained with a different held out speed upon which it is tested.

A reference for each of the 12 model types is provided in Table 5.2. Each model name is prefixed with the tactile channel used. STC/DTC refers to whether the model is for 'static texture classification' or 'dynamic texture classification'. $\boldsymbol{v}$ signifies that *leave-one-speed-out cross validation* is used. In this case, 10 models are constructed and $\boldsymbol{v}$ refers to the held out speed.

5.4 Perception: Static Models of Touch

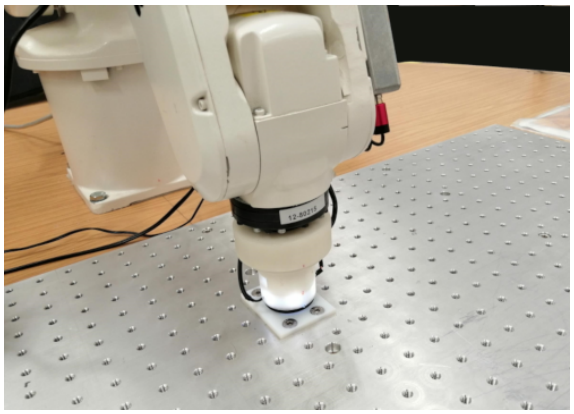

Figure 5.1: The TacTip is pressed onto a flat textured stimulus.

As outlined in Section 2.1.3.6, evidence suggests that perceptual cues of coarse textures may be mediated by static touch [57] and are encoded spatially within firing rates of populations of SA-I and FA afferents [43, 110, 112, 118]. Here we aim to understand the capabilities of the modified TacTip for texture perception via a purely spatial encoding in devised artificial SA-I and FA afferents (see Section 4.5.2).

As described in the previous chapter, Section 4.1, output of artificial SA-I and FA afferents are modelled on *firing rates* of individual SA-I and FA afferents; i.e., here we assume a *rate code* as the fundamental code for single afferents, upon which more abstract codes, such as a spatial code, may be composed. Described in Section 2.1.2 of the literature review, a *rate code* is one in which relevant stimulus information is

"correlated only with the average number of spikes within the encoding window" [73], where the *encoding window* is the duration of time that the spike train is used to convey information about the relevant stimulus feature. Therefore, in our case, the *encoding window* is analogous to the time over which each sample was collected.

Data collected via pressing (Section 5.2.2) and sliding (Section 5.2.3) are used in two separate experiments. The advantage of using data collected via sliding is that we can use the same dataset when comparing spatial encoding with spatio-temporal approaches to encoding. Data collected via pressing is also of interest since it may provide a stimulation which is indicative of a more natural explorative touch that might be employed with a purely spatial code.

5.4.1 Methods

5.4.1.1 SA-I Models of Artificial Static Touch

Sample Partitioning: Pressing

Pressing data is collected as described in Section 5.2.2.

For each stimulus class, samples are randomly separated into distinct training, validation and test sets of $N_{\text{train}} = 750$, $N_{\text{val}} = 50$ and $N_{\text{test}} = 50$ samples per class respectively, yielding total values of $N_{\text{tex}}N_{\text{train}} = 8450$, $N_{\text{tex}}N_{\text{val}} = 650$ and $N_{\text{tex}}N_{\text{test}} = 650$ samples respectively.

Sample Partitioning: Sliding

Sliding data is collected as described in Section 5.2.3.

The data for each stimulus condition is truncated to 5000 frames (~55 s) to ensure an equal number of samples per class. Samples are generated from individual frames.

Data is separated into train, validation and test sets using two distinct methods named ***All speed testing*** and ***Leave-one-speed-out cross validation***

(i) ***All speed testing:*** Within each stimulus class (textures), data collected at all speeds are grouped together to create $N_{\text{samples}} = 50000$ samples per class. For each stimulus class, samples are separated into distinct training, validation and tests sets, of $N_{\text{train}} = 45000$, $N_{\text{val}} = 2500$ and $N_{\text{test}} = 2500$ samples per class respectively, yielding total values of $N_{\text{tex}}N_{\text{train}} = 585000$, $N_{\text{tex}}N_{\text{val}} = 32500$ and $N_{\text{tex}}N_{\text{test}} = 32500$ samples respectively. Importantly, rather than randomly separating training, validation and test sets, they are split chronologically; i.e., the first 45000 frames are used for training, the next 2500 are used for validation and the last 2500 are used for testing. This serves to increase the novelty between sets: If random splitting were used, it is likely that many test samples would be chronologically adjacent to a training sample, thus, these samples would appear nearly identical enabling easy classification of the respective test sample. Furthermore, the validation set serves as a break between training and test sets which further increases novelty of the test set.

(ii) ***Leave-one-speed-out cross validation:*** 10 sets are constructed from the original dataset. In each set a different stimulus speed is 'held out' and randomly split 50:50 for validation and test sets. The remaining 9 speeds are used for training. This approach yields the same number of train, validation and test samples per set as the *all speed testing* method.

The motivation for using these two methods was to gain an understanding of how sliding speed affects static tactile images. If images generated with sliding stimuli are a product of spatial characteristics alone, we would expect perceptual performance of the two approaches to be comparable. Similarly, using a held out speed for testing further increases its novelty as the rotating drum stimulator was paused between each new collection speed.

Feature Extraction: Pressing

Artificial SA-I afferents are derived using the methods described in Section 4.5.2. In essence, image processing is used to track markers and an algorithm computes the Euclidean distance of each marker from its at-rest position. These Euclidean distances are equated to SA-I firing and are used to construct SA-I tactile images on a frame-by-frame basis.

Each sample is made from the last frame of each press; i.e., the lowest point.

Feature Extraction: Sliding

For each frame, features are derived as with pressing. In this case, however, each sample is a single frame.

Model

The SA-I static texture classification models (SA1-STC-*press*/SA1-STC-*slide*) are constructed with a 2D convolutional neural network, built using Keras [286], with TensorFlow backend [287]. The same network architecture is used for static perception with both pressing and sliding data, a schematic for which is shown in Fig. 5.2. This network permits 19x19 tactile images as input and outputs a 13x1 vector with each element corresponding to an individual texture class. All hidden layers use ReLU activation functions and the output layer uses softmax activations on each neuron.

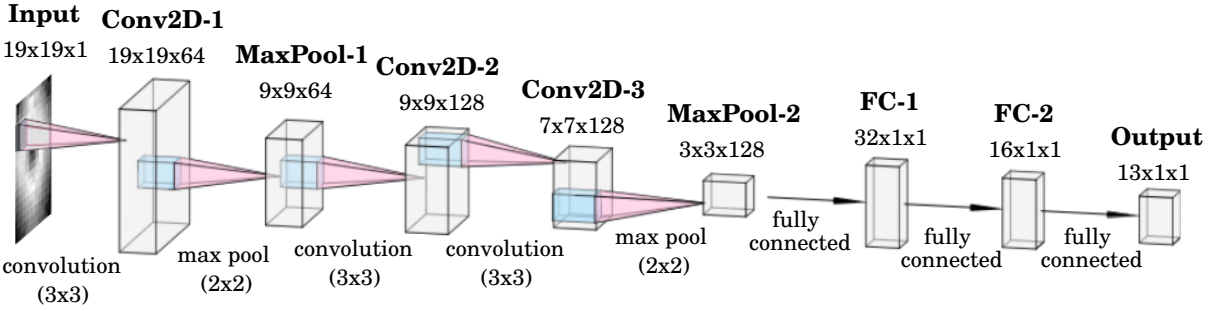

Figure 5.2: 2D-CNN architecture, used for texture classification using tactile images of artificial SA-I afferents (figure created with NN-SVG [285]).

In addition to the architecture shown in Fig. 5.2, regularization techniques are implemented: drop-outs of 0.4, 0.2 and 0.2 are used prior to layers FC-1, FC-2 and Output respectively as well as L2 regularisation (with factor 0.005) on each dense layer. Batch normalisation is implemented after each convolutional layer.

This architecture was chosen by manually tuning hyper-parameters and optimising for accuracy on the validation data.

Training

SA1-STC-*press* and SA1-STC-*slide* models are trained in Keras with TensorFlow backend. Training data is fed through the networks in batches of 64 samples. The entire set is propagated for a maximum of 150 epochs. An Adaptive Momentum Estimation optimiser (ADAM) on a categorical cross-entropy loss is used for updating the weights after each batch. Training may stop early if validation accuracy plateaus with the use of a patience factor of 30 epochs. The models are saved after every epoch. When performing *leave-one-speed-out cross validation*, a separate model, SA1-STC-*slide*-$\boldsymbol{v}$, is trained for each held out speed, v. A total of 12 models are trained, 1 for pressing data, 1 for *all speed testing* and 10 for *leave-one-speed out cross validation*.

Testing

The models achieving the highest validation accuracy are used for testing. Testing is performed by making predictions on the hold-out (test) sets. Test samples are fed through the networks and predictions are made by performing an argmax function on the network's outputs.

5.4.1.2 FA Models of Artificial Static Touch

Sample Partitioning: Pressing

The same pressing dataset of TacTip video recording is used for both artificial SA-I and FA afferents. Artificial FA samples are partitioned in an identical fashion to their SA-I counterparts.

Sample Partitioning: Sliding

Again, the same sliding dataset of TacTip video recording is used for both artificial SA-I and FA afferents. Artificial FA samples are partitioned in an identical fashion to their SA-I counterparts.

Feature Extraction: Presses

Artificial FA afferents are derived using the methods described in Section 4.5.2. As with SA-I afferents, marker positions extracted using image processing are used. In essence, the difference in marker displacement between adjacent frames (speed) is used to construct a tactile image of artificial FA firing.

FA samples are constructed from the mean of each afferent over the course of the entire press. Artificial FAs only responds to changes in marker position (see Section 4.4.2), therefore firing must be accumulated over the dynamic phase of a press.

Feature Extraction: Sliding

For each frame, features are derived as with pressing. In this case, however, each sample is a single frame.

Model

The FA static texture classification models for pressing and sliding (FA-STC-*press*/FA-STC-*slide*) are constructed from the same 2D convolutional architecture as with artificial SA-I afferents (Fig. 5.2).

Training

FA-STC models are trained in the same way as SA1-STC models (Section 5.4.1.1).

Testing

FA-STC models are tested in the same way as SA1-STC models (Section 5.4.1.1).

5.4.2 Results

5.4.2.1 Pressing

Data: Visual Inspection

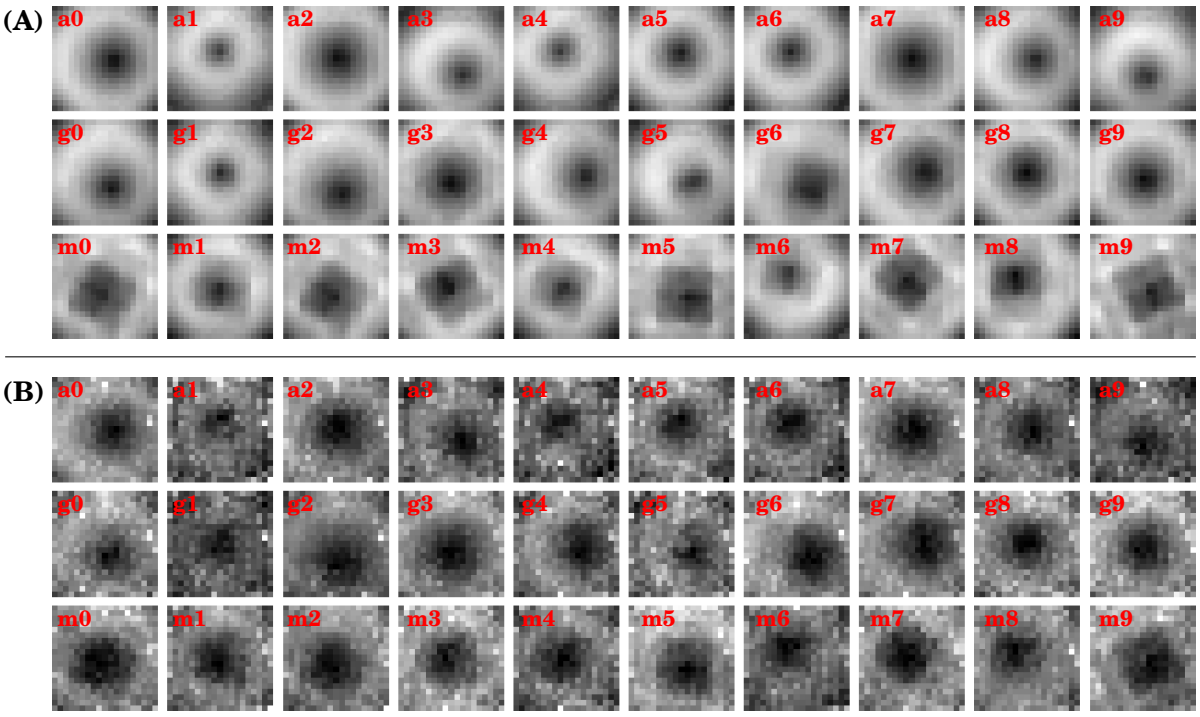

Figure 5.3: Tactile images of artificial SA-I and FA firing, **A** and **B** respectively, collected with presses. Light pixels indicate higher relative artificial firing. Letters, ***a***, ***g*** and ***m***, indicate stimulus class as described in Table. 5.1. Numbers (0-9) indicate sample instances.

Figs. 5.4A and 5.4B provide examples of tactile image samples of artificial SA-I and FA firing respectively, when the TacTip was pressed onto stimuli ***a***, ***g*** and ***m*** (described in Table 5.2). Nine samples are shown for each stimulus, where each sample was collected with x, y, z, roll, pitch and yaw varied randomly as explained, in Section 5.2.2.

Considering, first, SA-I tactile images (Fig. 5.3A), collected on stimulus ***a***, the smooth stimulus. In general, the form of these tactile images can be interpreted in much the same way as we did for the *Data Visualisation* section (Section 4.5.3.1) in Chapter 4. Essentially, as an area of the skin conforms to the flat surface, the amount of afferent firing increases away from the centre of the conformed area (see Fig. 4.10). Although broadly central, the position of the region of lower firing shifts between samples. This we attribute to $\pm 3.6°$ jitter in roll and pitch. Also, the size of this region changes which we attribute to ± 0.7 mm jitter in z. This interpretation seems to apply for stimulus ***g*** and to some degree stimulus ***m*** also.

Visually, we do not observe any obvious features of the tactile image which robustly signify texture class. In particular, classes ***a*** and ***g*** qualitatively appear to exhibit comparable intraclass and interclass variance. This is consistent with observations from our *Comparison with Physiological Data* of artificial SA-I SRPs produced on aperiodic gratings (Section 4.6.3.1, Chapter 4) on the 'effects of neighbouring ridges'. In particular, that narrow ridges or edges were indistinguishable from flat surfaces when spaced closer than ~4 mm apart (Fig. 4.13); i.e., where raised bumps are separated by <~4 mm, they will feel the same as a flat surface.

Class ***m***, which is the roughest of the 13 stimuli, however, does appear to show some consistent visual features which could distinguish these images from the others. This could be described as a more angular rectangular patch of low firing in the centre of the image. Class ***m*** has bumps which are separated by 6 mm, which is above the approximate spatial threshold observed from SRPs (~4 mm) (Section 4.6.3.1, Chapter 4). Therefore, the distinguishability of tactile images collected on class ***m*** from those of class ***a*** may serve as evidence for the interpretation from SRPs of the limits on spatial resolution. In this case, we would expect that classes ***a***-***i*** are not easily discriminated using from spatial codes of these images. Whereas, classification performance may increase for classes ***j*** onwards.

FA tactile images shown in Fig. 5.3B were from samples corresponding to those shown in Fig. 5.3A. Each FA tactile images exhibits similar overall structure to its corresponding SA-I image; e.g., a central dark patch of similar size, shape and position within the image as its corresponding SA-I image. We observe increased noise in FA images as compared with their SA-I counterparts.

These FA images were constructed from integrating FA firing over an entire tap. Therefore, it is perhaps unsurprising that the signal within SA-I and FA images looks similar considering that artificial FA firing is essentially the magnitude of the differential of SA-I firing. It may also follow that the increased noise is a consequence of negating the directional component of marker velocities for modelling FA firing; e.g., here marker velocities are summed regardless of whether the marker is moving further away or closer to its at-rest position, therefore integrating noise.

Texture Classification

Prediction results for models SA1-STC-*press* and FA-STC-*press* (models described in Table. 5.2) are shown as confusion matrices in Figs. 5.4A and 5.4B respectively. Table 5.3 provides accuracies and standard deviations for SA1-STC-*press* and FA-STC-*press* models.

Despite the apparent ambiguity in texture class based on a visual inspection of SA-I data (Fig. 5.3A), the CNN trained with SA-I tactile images, SA1-STC-*press*, was able to accurately predict texture class with moderate consistency. The modal predicted class was always the true class and the variance in prediction, averaged across classes, was moderate: ~1.2 classes. Overall prediction accuracy was ~72% correct. Therefore, our hypothesis, based on tactile images and spatial response profiles SRPs on aperiodic gratings (Section 4.6.3.1, Chapter 4), that the model would perform poorly when discriminate classes ***a-i*** turned out to be incorrect. It is not obvious why our hypothesis did not come to fruition but clearly the model was able to detect some robust feature that is not immediately discernable from the tactile images. It may be possible that some non-intended cue might have been available during data collection, such as variation in stimulus height between changing stimulus class, although, a jitter of ±0.7 mm in z would seem larger than any likely variation in 3D-printing thickness.

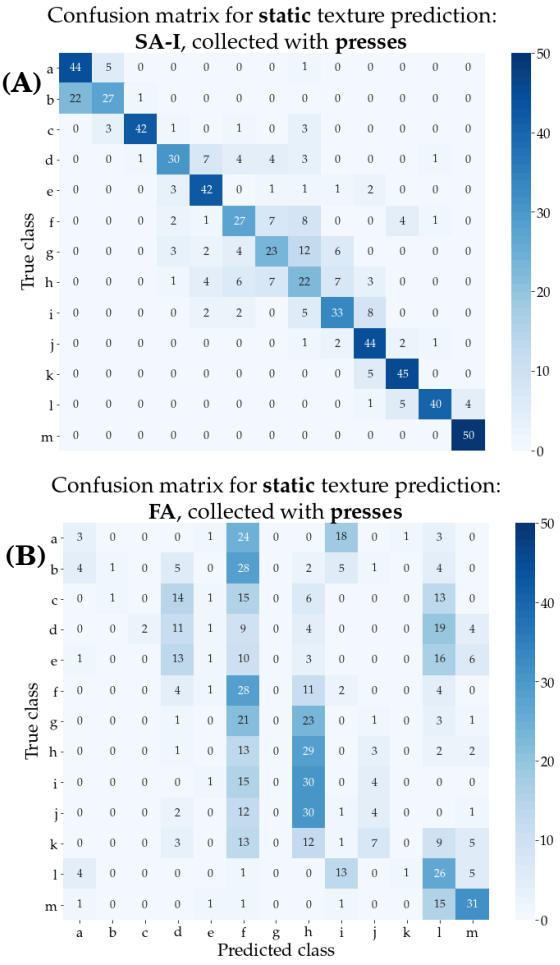

Figure 5.4: Confusion matrices of static texture prediction for data collected with presses. **A** and **B** show results for models SA1-STC-*press* and FA-STC-*slide* trained with SA-I and FA respectively.

There is an apparent trend, albeit loose, for better predictions towards the roughest and smoothest stimuli; i.e., at the edge of the range of stimuli. This may be because textures located adjacently are the most similar in terms of dimensionality, thus classes at either end of the roughness scale have are fewer similar stimuli making prediction easier.

In order to understand the degree to which spatial features available to coarse stimuli,

j*-*m, may have aided perception, we compare prediction accuracy and variance on these four classes, with the same metrics on the 4 smoothest stimuli, ***a*-*d***. We compare extremes of the range because both extremes benefit from the effect described above: essentially having fewer similar alternatives; i.e., it would be unfair to compare results on classes ***j*-*m*** with the remaining stimuli because, on average, these 4 stimuli have fewer similar alternatives than the remaining 9. Prediction accuracy on stimuli ***j*-*m*** was ~90% compared with ~72% on stimuli ***a*-*d*** and the variance averaged across classes on stimuli ***j*-*m*** was ~0.2 classes compared with ~1.4 classes on stimuli ***a*-*d***. This suggests that, despite the unpredicted levels of performance for stimuli ***a*-*i***, there is perhaps some attenuation owing to limits on spatial resolution of the artificial SA-I afferents.

In contrast, texture prediction using artificial FA afferents with presses (model FA-STC-*press*) was poor (~21 % accuracy and ~6.7 classes variance), despite the visual similarity of the data to SA-I (Fig. 5.3). The variance was relatively low considering the poor performance of the model but this is because of bias; e.g., for classes ***f*** and ***h*** (see Fig. 5.4). An explanation for this drop in performance is the increased noise observed with artificial FA firing compared with SA-I: With increased noise, the signal becomes 'washed-out'. As a consequence, during training, model weights are less responsive to features which might accurately inform texture.

5.4.2.2 Sliding

Data: Visual Inspection

Examples of tactile images generated with sliding are shown in Figs. 5.5A and 5.5B and Figs. 5.5C and 5.5D for artificial SA-I and artificial FA afferents respectively. As with presses, Figs. 5.5A and 5.5C show examples collected with stimulus classes ***a***, ***g*** and ***m*** (described in Table 5.2) and 10 examples are shown for each stimulus. In this case, each example within a given class was collected at a different speed (10-100 mms^{-1}), indicated by the image label.

There are clear differences in the general characteristics of these tactile images when comparing to pressing images (Figs. 5.3A and B). For example, they are in general darker, which means less overall firing; i.e., the pins did not deflect as far from their rest positions when compared with pressing.

Images of SA-I firing (Figs. 5.5A and Fig. 5.5B) tend to exhibit more asymmetry than those collected with presses. By this, we mean that rather than a central dark patch of low firing surrounded by a lighter area, these images often show a lighter crescent shape to the left or right of the image (e.g., **g-40**, **g-30**, **m-20** etc. in Fig. 5.5A). Naively, it might be expected that sliding tactile images would look more similar to those collected with presses since, in both cases, at any given frame, the TacTip was in contact with the same stimulus. These results, however, suggest that when stimulated via sliding, the TacTip's deformation did not reflect spatial characteristics of the stimulus but rather was dominated by shear effects.

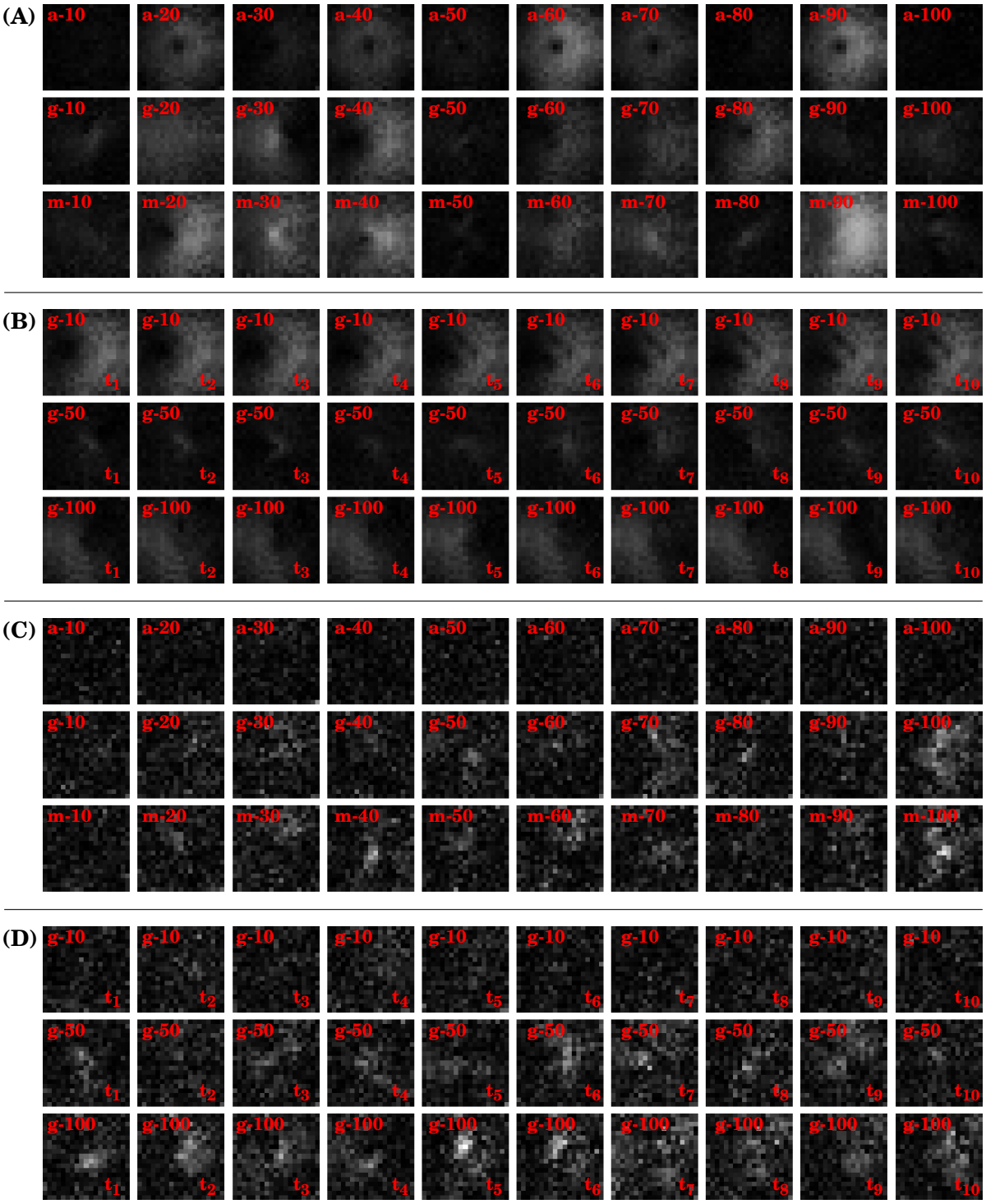

Figure 5.5: Tactile images of artificial SA-I (**A** and **B**) and FA (**C** and **D**) firing collected with sliding touch. Labels indicate stimulus (letter) and speed (number). **A** and **C** show a sample from every speed on stimuli ***a***, ***g*** and ***m*** (described in Table 5.2). **B** and **D** show time series of 10 adjacent samples, collected at speeds 10, 50 and 100 mms^{-1}, all on stimulus ***g***. The label at the bottom right of each image in **B** and **D** indicate the temporal ordering of each frame.

Similar to pressing data, we do not observe any features in SA-I images which can robustly signify texture class. For example, images **m-20** and **g-40** in Fig. 5.5A look qualitatively similar, whilst the variation within classes ***m*** and ***g*** is significant. We suspect this was likely caused by the varying speed. For reference, Fig. 5.5B shows tactile images of artificial SA-I on stimulus class ***g*** (class with the median bump spacing). Rows are separated according to speed (10, 50 and 100 mms^{-1}) and consist of time series of 10 adjacent frames (increasing time from left to right). Images appear similar along rows and comparatively dissimilar down the columns. Considering Figs. 5.5A and 5.5B, we conclude that the spatial modulation of artificial SA-I firing over the TacTip is highly dependent on both texture and speed.

Tactile images of artificial FA firing shown in Figs. 5.5C and 5.5D. are made from frames corresponding to those shown in Fig. 5.5A and Fig. 5.5B respectively.

As with artificial SA-I, there are clear differences in the general characteristics of these tactile images when comparing to images created with presses (Fig. 5.3). In particular, considering FA images in isolation, there is considerably less structure with data collected through sliding compared with pressing. This is likely a consequence of, not just the nature of stimulation, but also a difference in the way samples were constructed: For example, with sliding data, each FA tactile image is effectively constructed from the difference in adjacent SA-I images. Considering Fig. 5.5B, adjacent SA-I images appear very similar, thus any structure in the associated FA image is likely to be minimal. In comparison, FA samples produced with pressing (Fig. 5.3B) were constructed from the accumulation of FA firing over the course of a each press. By integrating small signals at each frame, the overall signal to noise ratio will be increased.

As with artificial SA-I samples, FA images collected with sliding touch do not exhibit any clear visual features which robustly signify texture class. Contrastingly, however, artificial FA firing does not show the same degree of consistency between adjacent images (Fig. 5.5D).

Texture Classification - *All speed testing*

Prediction results for models SA1-STC-*slide* and FA-STC-*slide* (models described in Table. 5.2) are shown as confusion matrices in Figs. 5.6A and 5.6B respectively. Table 5.3 provides accuracies and standard deviations for SA1-STC-*slide* and FA-STC-*slide* models.

The trained model, SA1-STC-*slide*, was able to correctly predict texture in all but 4 trials. This model was trained with samples collected at all 10 speeds (*all speed testing*). Meaning that, for each test trial, the exact stimulus condition shown to the model had also been used to train the network. Although the tactile images of SA-I firing (Fig. 5.5A) appear to show significant variance within each class, the high accuracy is consistent with observations from Fig. 5.5B that tactile data collected under identical stimulus conditions exhibits minimal variance.

FA-STC-*slide*, performed comparatively worse than its SA-I counterpart (SA1-STC-*slide*), however, dramatically outperformed its equivalent model for pressing data (FA-STC-*press*). The overall prediction accuracy was ~58% correct. The modal predicted class was always the true class. The variance in prediction, averaged across classes, was relatively high: ~7.8 classes. It

seems likely that the drop in performance, as compared with the equivalent model trained on SA-I data (SA1-STC-*slide*), is a consequence of the relative inconsistency in FA tactile images collected at the same speed; e.g., more variance is seen within the rows of Fig. 5.5D compared with Fig. 5.5B. Owing to this, unlike model SA1-STC-*slide*, FA-STC-*slide* does not benefit so much from seeing examples of each speed at both train and test time.

Texture Classification - *Leave-one-speed-out*

Prediction results for *leave-one-speed-out cross validation* models SA1-STC-*slide*-***v*** and FA-STC-*slide*-***v*** (models described in Table. 5.2) are shown as confusion matrices in Figs. 5.7A and 5.7B respectively. Table 5.3 provides accuracies and standard deviations for SA1-STC-*slide*-***v*** and FA-STC-*slide*-***v*** models averaged across hold out speeds.

Firstly, considering models SA1-STC-*slide*-***v***, trained with SA-I data, clearly performance was significantly worse than in the case of *all speed testing* (SA1-STC-*slide*, Fig. 5.6A) which can be understood as a consequence of the variance in tactile images across speeds. Despite this, each model clearly performed better than chance which suggests that there is at least some level of coherence within texture classes in terms of speed-invariant spatial features. The accuracy averaged across all models was ~50 %. Near perfect predictions were made for stimulus class ***a*** (completely smooth) at every hold-out speed and we see a general trend towards better performance at either end of the range of stimuli. As with pressing data, this may be understood as a consequence of fewer similar alternatives at the range edges.

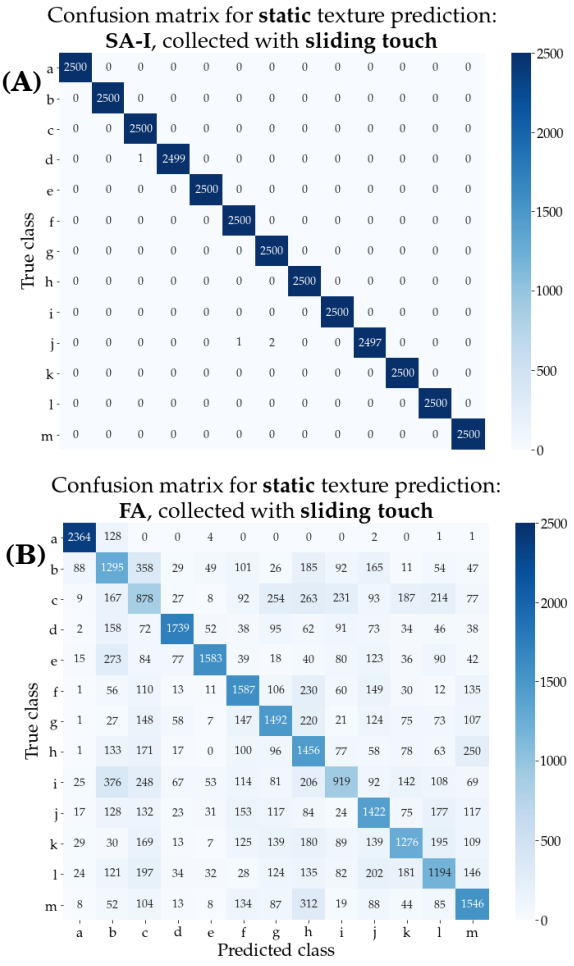

Figure 5.6: Confusion matrices of static texture prediction for *all speed testing*. **A** and **B** show results for SA1-STC-*slide* and FA-STC-*slide* trained with SA-I and FA afferents respectively.

The best performing SA-I model was SA1-STC-*slide*-80, with a hold-out speed of 80 mms^{-1}, achieving a classification accuracy of ~64 %. It seems likely that performance would be worse on

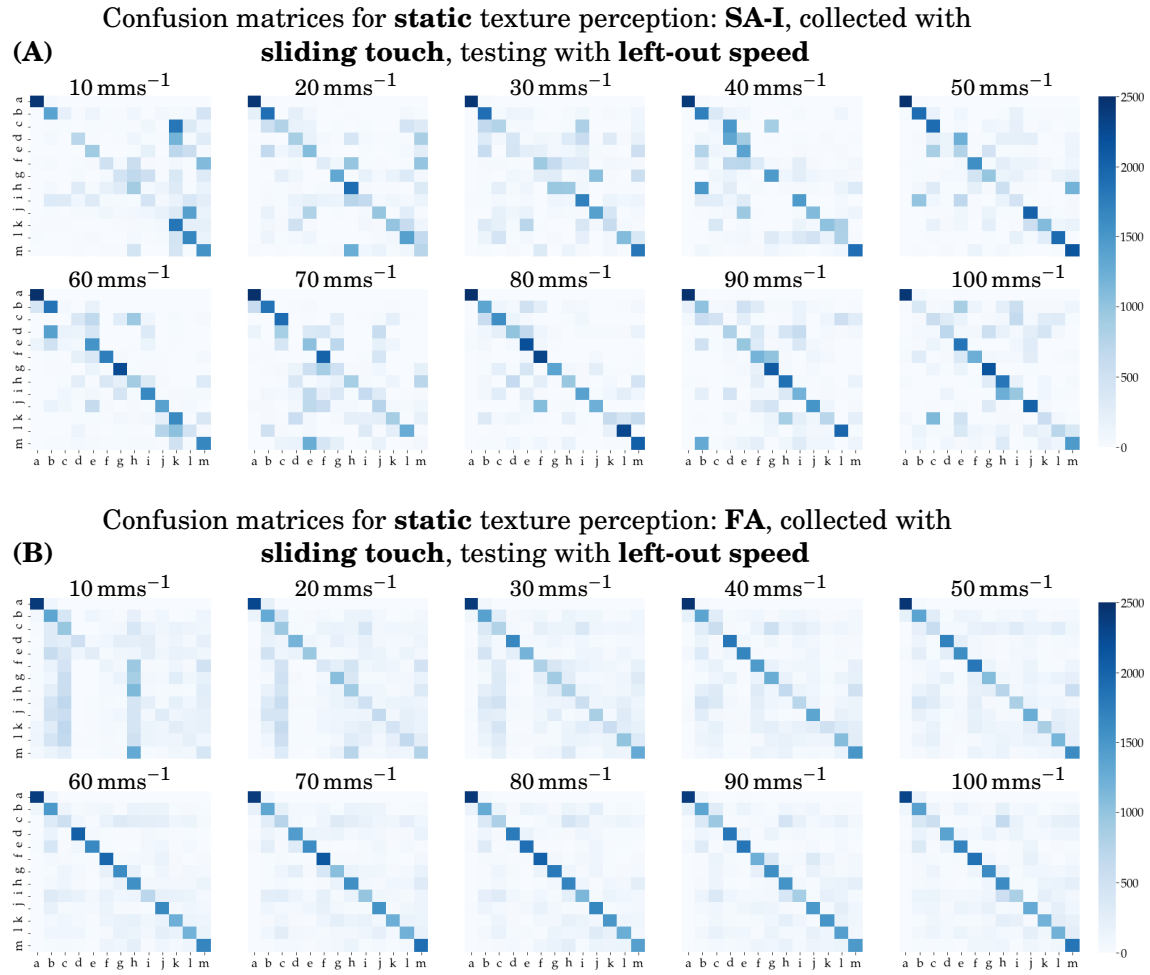

Figure 5.7: Confusion matrices of static texture prediction using *leave-one-speed-out cross validation*. **A** and **B** are for artificial models SA1-STC-*slide*-$\boldsymbol{v}$ trained on SA-I and FA respectively. Each matrix refers to a different model trained and tested with a held out speed, indicated by its individual title.

hold-out speeds at the extremes of the speed range because the network was trained only on data collected either at faster or slower speeds for 10 and 100 mms^{-1} respectively. This might have a continued but lessening affect when moving towards the middle of the range of hold-out speeds. Whilst this hypothesis might explain the relatively poor performance for the slower hold-out speeds, there does not appear to be significant attenuation in performance at the other end of the range. This could be attributed to a competing relationship between hold-out speed and performance; e.g., more salient data being created when the stimulation speed is increased.

The average accuracy of static FA models, FA-DTC-$\boldsymbol{v}$, across held out speeds was ~51 %. This does not seem to reflect the comparative performance to equivalent SA-I models as seen from confusion matrices, where, in general, FA models (Fig. 5.7B) appear to perform better than SA-I models (Fig. 5.7A). This is because, in fact, FA-DTC-10 had considerably worse accuracy (~26 %)

for hold-out speed 10 mms^{-1} than SA-DTC-10 (~40 %), which reduced the average accuracy score for FA-DTC-$\boldsymbol{v}$ models.

FA-DTC-$\boldsymbol{v}$ models provided better accuracy for hold-out speeds v >60 mms^{-1} than the FA model trained and tested using all speeds (FA-STC-*slide*, Fig. 5.6B), forming an interesting contrast to that seen with artificial SA-I afferents.

FA shows a similar trend to SA-I, with a reduction in performance for slower hold-out speeds, suggesting that, within the range used here, perception is improved with higher sliding speed. This argument also offers explanation for the improvement in performance over *all speed testing* at hold-out speeds greater than the 60 mms^{-1} threshold: If prediction accuracy is poor for slow speeds, by removing these speeds when testing, accuracy may improve, perhaps above a baseline level achieved with *all speed testing*. This hypothesis was confirmed by simply removing test samples collected at speeds below the 60 mms^{-1} threshold and re-testing with the *all speed* model (FA-STC-*slide*) which achieved an accuracy of ~80 %.

The fact that FA afferents were generally better predictors of texture than SA-I afferents when using a held out speed, suggests that artificial FA afferents offer better generalisation to previously un-observed speeds. Or, equivalently, SA-I models tend to over-fit, responding to spatial structures which vary with speed, rather than speed-invariant features. This hypothesis is supported by comparing Fig. 5.5B of SA-I firing with 5.5D of FA firing. Unlike with SA-I tactile images, there appears to be little coherence, in terms of spatial structure, within speeds for FA firing. Thus, FA data offers less opportunity for the model to learn undesirable features; i.e., the data is naturally regularised.

5.4.3 Discussion

Spatial encoding of texture was isolated in all three experiments: pressing, sliding - *all speed testing* and sliding - *leave-one-speed-out* by removing the temporal dimension to each sample and designing the stimuli to have consistent intensive qualities (see Section 5.2.1.1).

In all cases, texture prediction with artificial SA-I afferents was considerably better than chance, suggesting that spatial modulation of the modelled SA-I channel can encode roughness. Similarly, artificial FA tactile images were found to be good predictors of texture in all but pressing scenarios.

When the TacTip was pressed onto texture stimuli, no robust visual feature for roughness encoding was immediately obvious in either SA-I or FA firing, however, model SA1-STC-*press*, trained with SA-I data, which exhibits a spatial feature detector in the form of convolutional layers, was able to predict texture with moderate accuracy (~72 %), suggesting there was indeed some spatial feature which robustly varied with texture. This supports the likeness of our afferent model to the human SA-I channel, since evidence suggests that coarse textures may be encoded in the spatial modulation of activity within populations of natural SA-I afferents [43, 110, 112, 118]. Furthermore, comparison of prediction accuracy on coarse stimuli with that on fine stimuli

suggests a modulation in performance which agrees with our observations of spatial resolution from spatial response profiles (SRPs) collected on aperiodic gratings (Section 4.6.3.1, Chapter 4).

Unlike the results of human studies, which generally conclude that FA afferents may also provide feasible spatial coding of coarse textures, we found that a CNN trained with artificial FA tactile images collected with presses (FA-STC-*press*) was unable to classify texture. A result which we attribute to increased noise resulting from the approach to generating FA tactile images with presses (Section 5.4.2.1).

Clearly sliding data for artificial SA-I afferents varied with both roughness and speed (Figs. 5.5A and 5.5B). The tactile images seem to be characteristic of shearing. Despite this, accurate predictions of texture were made, in particular with *all speed testing*, SA1-STC-*slide*, suggesting that, as well as spatial, frictional properties may provide a valid cue for roughness.

Table 5.3: Table of results for all static texture classification models. Variance here is the variance of prediction within a class, averaged accross all classes. Where the models are trained and tested with *leave-one-speed-out cross validation* (indicated with $\boldsymbol{v}$), the results are averaged across all left-out speeds.

Model	Accuracy (%)	Variance (classes)
SA1-STC-*press*	72	1.2
FA-STC-*press*	21	6.7
SA1-STC-*slide*	100	0.0
FA-STC-*slide*	58	7.8
SA1-STC-*slide*-$\boldsymbol{v}$	50	6.4
FA-STC-*slide*-$\boldsymbol{v}$	51	8.3

SA1-STC-*slide*-$\boldsymbol{v}$ performed worse than SA1-STC-*slide*, which is most likely a consequence of the variation in tactile images with speed. This again, indicates that deformation was dominated by frictional rather than spatial properties since purely spatial cues will be unaffected by scanning speed.

Unlike SA-I models, those trained and tested with artificial FA afferents generalised much better to previously unobserved speeds (Fig. 5.7), an interesting result which we attribute to implicit regularisation within FA training data (Section 5.4.2.2: Texture Classification - *Leave-one-speed-out*).

The improvement of both SA and FA STC-*slide* models on their pressing counterparts may be a consequence of multiple factors: Frictional cues provided via sliding touch may be a better predictor of texture than spatial cues provided by pressing. However, a considerably larger quantity of data was used to train sliding models, which is known to improve generalisation. It seems likely that the later hypothesis would have contributed significantly to the improvement in performance, and consequently, makes it difficult to assess the feasibility of the former.

As with the pressing method, neither *all speed testing* nor *leave-one-speed-out* models showed significant noticeable degradation in performance with increasing fineness of the stimulus. It is possible that, as with the pressing method, this is again a consequence of some non-intended cue becoming implicitly available within the tactile image data as a result of the rigid experimental

setup. We suggest, however, that in the case of sliding data the unusually high performance seen across the stimulus range can be predicted from the tactile image data, which, as stated, is dominated by frictional rather than spatial cues. We believe valid frictional cues are likely to be available for the finest stimuli were spatial cues (based on stimulus properties) are not.

5.5 Perception: Dynamic Models of Touch

Whilst it is believed that humans employ spatial coding schemes for perceiving coarse textures, it is thought that below a threshold coarseness, textures can no longer be spatially resolved by SA-I and FA afferents [221]. At this point vibrations created by dynamic touch provide cues for surface roughness which can be temporally encoded [126, 129] (Section 2.1.3.6).

We therefore consider dynamic touch as a method for robot texture perception and devise methods which leverage a spatio-temporal coding of artificial SA-I and FA firing and temporal cues of acoustic vibrations measured with the TacTip's vibrational channel.

As mentioned in the previous chapter (Section 4.1), unlike neuromorphic sensors and, indeed, neurons, our artificial afferents do not produce spiking output, instead, we suggest each afferent models *firing rate*, a quantity which is often implied as a fundamental coding scheme for individual neurons in biology, in contrast many neuorscientists have argued that precise inter-spike timing encodes important information [266]. In this sense, when we talk about temporal code of artificial SA-I and FA afferents, we do not imply that this is analogues to schemes that use precise timing of individual spikes, but rather use temporal variation in *firing rate* within an encoding window, which is also considered temporal coding [74].

Here we only utilise data collected via sliding (Section 5.2.3) since pressing data does not provide time varying signals. Data collected via sliding is indicative of natural texture exploration employed by humans [288].

5.5.1 Methods

5.5.1.1 SA-I Models of Artificial Dynamic Touch

Sample Partitioning

The same sliding dataset is used for dynamic touch as with static touch. For details of how this set is collected see Section 5.2.3.

As with static touch, the data for each stimulus condition is truncated to 5000 frames (~55 s) to ensure an equal number of samples per class. In contrast to static touch, samples are generated from sets of 10 adjacent frames (~0.1 s), this is to enable a temporal encoding of texture, yielding a total of 500 samples per stimulus condition. In the analogy with human perception, this would be a ~0.1 s encoding window. Importantly, the temporal dimension of each sample is not reduced in anyway; i.e., each sample is now a 19x19x10 3D vector, where the decoding model is able to extract features within this full 3D spatio-temporal structure.

Data is separated into train, validation and test sets using the same distinct methods as with static touch, collected with sliding stimulation: ***All speed testing*** and ***Leave-one-speed-out cross validation***. For details of how these two approaches differ see Section 5.4.1.

(i) ***All speed testing:*** In dynamic touch, the same ratio train:validation:test as with static touch is used. Thus, there are N_{samples} = 5000 samples per class in total and N_{train} = 4500, N_{val} = 250 and N_{test} = 250 samples per class respectively, yielding total values of $N_{\text{tex}}N_{\text{train}}$ = 58500, $N_{\text{tex}}N_{\text{val}}$ = 3250 and $N_{\text{tex}}N_{\text{test}}$ = 3250 samples respectively. As with static touch, rather than randomly separating training, validation and test sets, they are split chronologically in order to increase novelty between train, validation and test sets.

(ii) ***Leave-one-speed-out cross validation:*** 10 distinct sets are constructed from the original dataset. In each set a different stimulus speed is 'held out' and randomly split 50:50 for validation and test sets. The remaining 9 speeds are used for training. This approach yields the same number of train, validation and test samples per set as the *all speed testing* method.

The motivation for using these two methods differs from that of static touch: It is known from psychophysical experiments that human texture perception is invariant to speed [120, 130, 131]. It is therefore of interest to assess the degree to which our methods of texture perception are speed-invariant with the hope that we may gain insight into how this phenomenon is achieved in humans.

Feature Extraction

Artificial SA-I afferents of each frame are derived using the method described in Section 4.5.2. Thus each sample consists of a 10x19x19 feature vector of artificial SA-I firing.

Model

The SA-I dynamic texture classification model (SA1-DTC) is constructed with a convolutional-LSTM (ConvLSTM), built using Keras [286], with TensorFlow backend [287], a schematic for which is shown in Fig. 5.8. This network permits 19x19x10 tactile images as input and outputs a 13x1 vector with each element corresponding to an individual texture class.

In essence, a frame at each timestamp, $t_{0\text{-}9}$, of a sample is passed through a feature detector, labelled 'CNN'. The architecture of the feature detector is taken directly from that used in the STTCM model (Fig. 5.2: layers Conv2D-1 to MaxPool-2). The output feature maps are flattened before being passed to a multi-layer LSTM unit, consisting of 3 layers with 10 LSTM blocks each. The output of the LSTM unit is passed to a fully connected unit, 'FC', the architecture of which is taken directly from that used in the STTCM model (Fig. 5.2: layers FC-1 to FC-2). An output layer, consisting of 13 neurons, provides a prediction for each class.

Reusing the architecture of the feature detection and fully connected parts of the STTCM model serves to provide a more controlled comparison between encoding mechanisms; i.e., dif-

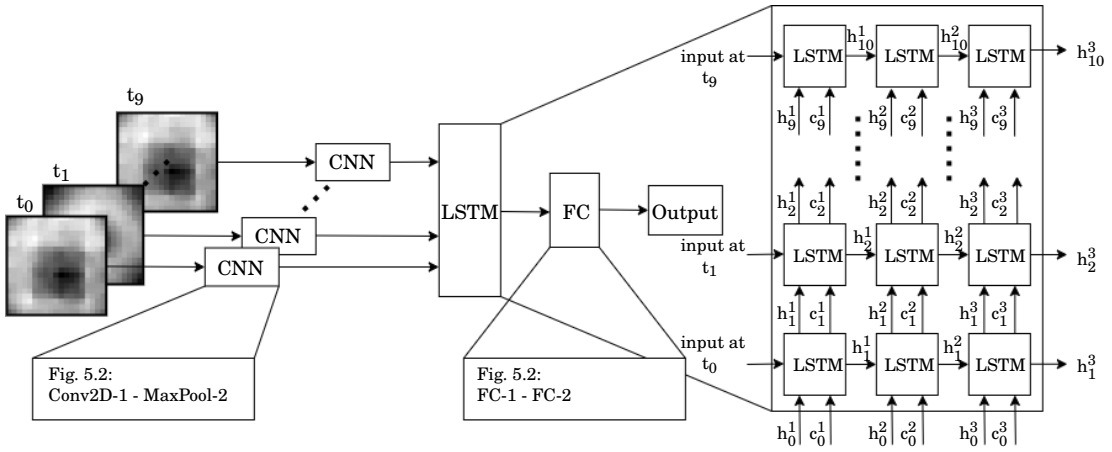

Figure 5.8: ConvLSTM architecture, used for texture classification with tactile image sequences of artificial SA-I and FA afferents.

ferences in performance between static and dynamic touch can more easily be attributed to the availability of temporal features. The LSTM hyper-parameters were chosen by manual tuning whilst optimising for validation accuracy.

Training

SA1-DTC models are trained in Keras with TensorFlow backend. Training data is fed through the networks in batches of 64 samples. The entire set is propagated for a maximum of 150 epochs. An Adaptive Momentum Estimation optimiser (ADAM) on a categorical cross-entropy loss is used for updating the weights after each batch. Training may stop early if validation accuracy plateaus with the use of a patience factor of 30 epochs. The models are saved after every epoch. When performing *leave-one-speed-out cross validation*, a separate model, SA1-DTC-$\boldsymbol{v}$, is trained for each held out speed, v. A total of 11 models are trained, 1 for *all speed testing* and 10 for *leave-one-speed out cross validation*.

Testing

In all cases the model achieving the highest validation accuracy is used for testing. Testing is performed by making predictions on the hold-out (test) set. Test samples are fed through the networks and predictions are made by performing an argmax function on the network output.

5.5.1.2 FA Models of Artificial Dynamic Touch

Sample Partitioning

The same sliding dataset of TacTip video recording is used to construct dynamic FA samples as is used for dynamic SA-I. Artificial FA samples are partitioned in an identical fashion to their SA

counterpart.

Feature Extraction

Artificial FA afferents of each frame are derived using the method described in Section 4.5.2. Thus each sample consists of a 10x19x19 feature vector of artificial FA firing.

Model

The FA dynamic texture classification models (FA-DTC) are constructed from the same ConvLSTM architecture as used with artificial SA-I afferents (Fig. 5.8).

Training

FA-DTC models are trained in the same way as SA1-DTC models (See Section 5.5.1.1).

Testing

FA-DTC models are tested in the same way as SA1-DTC models (See Section 5.5.1.1).

5.5.1.3 Vibration Model of Artificial Dynamic Touch

Sample Partitioning

Microphone data, collected simultaneously to the video used to construct dynamic SA-I and FA samples, is used to construct vibration samples. For details of how this set is collected see Section 5.2.3.

Overlapping samples of ~2 s and stride of ~1 s are generated from sets of 88200 data points, yielding a total of 58 samples per stimulus condition.

As with preceding modalities, data is separated into train, validation and test sets using two methods: ***All speed testing*** and ***Leave-one-speed-out cross validation***. For details of how these two approaches differ see Section 5.4.1.

(i) ***All speed testing:*** Approximately the same ratio train:validation:test as with artificial SA-I and FA afferents is used. There are N_{samples} = 580 samples per class in total and N_{train} = 520, N_{val} = 30 and N_{test} = 30 samples per class respectively, yielding total values of $N_{\text{tex}}N_{\text{train}}$ = 6760, $N_{\text{tex}}N_{\text{val}}$ = 390 and $N_{\text{tex}}N_{\text{test}}$ = 390 samples respectively. Again, rather than randomly separating training, validation and test sets, they are split chronologically in order to increase novelty between train, validation and test sets.

(ii) ***Leave-one-speed-out cross validation:*** 10 distinct sets are constructed from the original dataset. In each set a different stimulus speed is 'held out' and randomly split 29:30 for validation and test sets. The remaining 9 speeds are used for training. This approach yields $N_{\text{tex}}N_{\text{train}}$ = 6903, $N_{\text{tex}}N_{\text{val}}$ = 377 and $N_{\text{tex}}N_{\text{test}}$ = 390 samples respectively.

The difference in number of samples between *all speed testing* and *leave-one-speed-out cross validation* approaches is owing to a requirement that the same ratio of speeds be present in train, validation and test sets as with *all speed testing*, therefore a small portion of the data was thrown away. This is not the case for *leave-one-speed-out cross validation* and thus we were able to use all of the data for this method.

As with artificial SA-I and FA under dynamic stimulus conditions, the motivation for using these two sample partitioning techniques is to assess the degree to which our methods of texture perception are speed-invariant.

Feature Extraction

From each sample of raw audio data (time domain), a 1D feature vector of amplitudes at discrete frequencies is constructed by performing an FFT. Because each sample contains the same number of data points and spans the same amount of time, corresponding elements of the feature vectors of each sample are associated with the same frequencies. These feature vectors are truncated to a size of 400x1 which provides a maximum frequency of 200 Hz. The features are scaled to fall between 0 and 1 with a division by the largest amplitude: unscaled input values can cause slow or unstable learning [289]. This is important for the vibrational channel because the unscaled input values have a large range (Fig. 4.29)

Our hypothesis is that the harmonic structure of induced vibrations is a viable cue of texture (Section 4.7.3), therefore, each feature vector must be of sufficient resolution to communicate harmonic structure. There is, however, a trade-off between number of samples, N_{samples}, and detail within the feature vectors: FFT is a linear transformation that maps the N_{td} points to N_{fd} points, where N_{td} and N_{fd} are the number of time and frequency domain points respectively. Therefore, with more data-points per sample, each feature vector has higher resolution. However, with too few training samples, a predictive model is likely to over-fit when training. This trade-off was considered when deciding on how many samples should be constructed from the continuous dataset.

Model

The vibrational dynamic texture classification model (vibro-DTC) is constructed with a 1D convolutional neural network, built and trained using Keras [286], with TensorFlow backend [287]. A schematic for the chosen architecture is shown in Fig. 5.9. This network permits 400x1 feature vectors as input and outputs a 13x1 vector with each element corresponding to an individual texture class. All hidden layers use ReLU activation functions and the output layer uses softmax activations on each neuron.

In addition to the architecture shown in Fig. 5.9, regularization techniques are implemented: drop-out of 0.4, is used prior to layer FC-1, as well as L2 regularisation (with factor 0.005) on layer FC-1. Batch normalisation is implemented after each convolutional layer.

This architecture was chosen by manually tuning hyper-parameters and optimising for

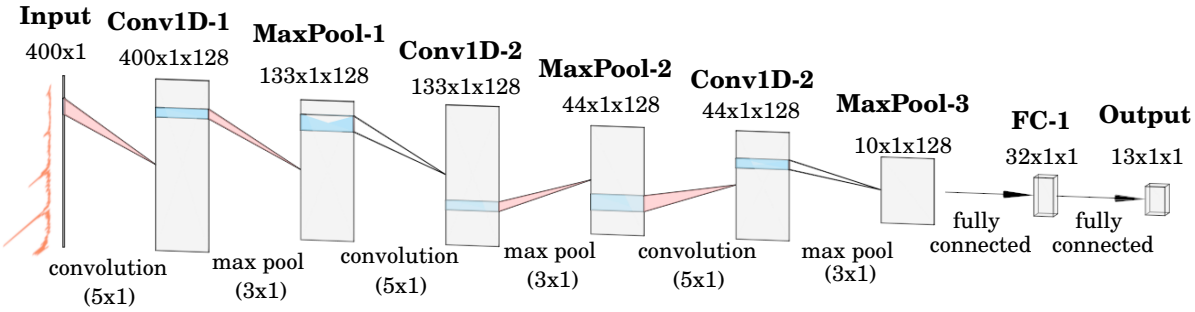

Figure 5.9: 1D-CNN architecture used for texture classification using FFTs of vibration data samples (figure created with NN-SVG [285]).

validation accuracy.

Training

Vibro-DTC models are trained in the same way as FA and SA1-DTC models (See Section 5.5.1.1). Again, a total of 11 models are trained, 1 for *all speed testing* and 10 for *leave-one-speed out cross validation*.

Testing

Vibro-DTC models are tested in the same way as FA and SA1-DTC models (See Section 5.5.1.1).

5.5.2 Results

Data: Visual Inspection

For dynamic perception experiments, we use the same data as with static touch collected through sliding contact. Therefore, as with static touch, here we refer to Fig. 5.5 for analysing SA-I and FA tactile images. In particular, we are concerned with Figs. 5.5B and 5.5D as these show sequences of 10 adjacent frames of SA-I and FA tactile images respectively; i.e., each row depicts an individual sample used with dynamic texture perception models, SA1-DTC and FA-DTC (described in table 5.2)

A striking characteristic of artificial SA-I afferents, evident from Fig. 5.5B, is that there is very little variation on the time scale presented (~0.1 s). This is compared to FA images (Fig. 5.5D), where we see considerably more temporal variation in the spatial modulation of artificial firing. This might indicate that artificial FA firing contains a more suitable temporal encoding of texture.

Examples of vibrational channel samples, generated as described in Section 5.5.1.3, are shown in Fig. 5.10. As described in Section 5.5.1.3, we hypothesise that the harmonic structure of induced vibrations; i.e., FFT shape, provides a viable cue for texture and therefore, each sample is required to exhibit sufficient resolution in order to accurately express the harmonic structure. Fig. 5.10 suggests that the chosen resolution provides sufficient detail to visually distinguish textures based on the shape of their FFTs. 200 Hz appears to be a suitable maximum frequency

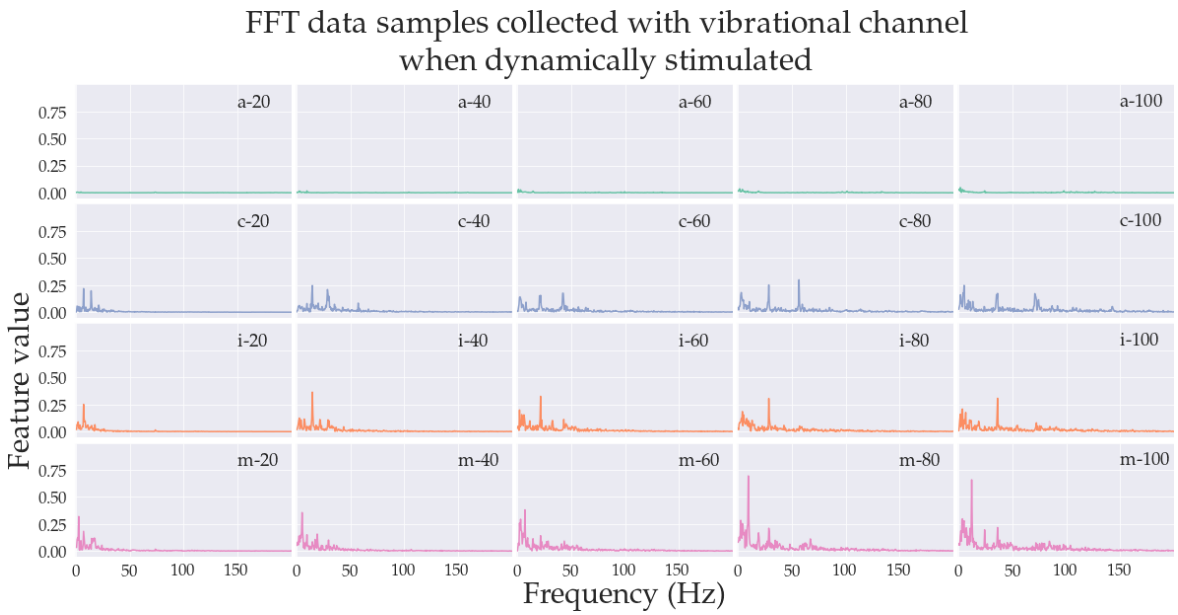

Figure 5.10: FFT samples produced when sliding on stimuli. Labels indicate stimulus (letter) and speed (number).

since we found there to be very little spectral power above this frequency in all cases, therefore, we used a vector length of 400 features, expressing frequencies up to 200 Hz.

The hypothesis that the harmonic structure of induced vibrations is a viable cue for texture discrimination is supported by the FFTs shown in Fig. 5.10: we observed a consistent shape within each texture, with speed only affecting the scaling rather than structure; e.g., at higher speeds the feature vectors are 'stretched-out' with peak frequencies increased and at lower speeds the feature vectors are 'squashed' with peak frequencies reduced.

On the scale of these plots, samples from class ***a*** (smooth stimulus) have extremely low feature values, which is a consequence of relatively low amplitude vibration. This is seen at all speeds and has the consequence of making samples of class ***a*** appear very distinguishable from those of other classes. We predict that all models trained with vibrational channel data will have high accuracy when tested on class ***a***.

Texture Classification - *All speed testing*

Prediction results for models SA1-DTC, FA-DTC and vibro-DTC (models described in Table. 5.2) are shown as confusion matrices in Figs. 5.11A, 5.11B and 5.11C respectively. Table 5.4 provides accuracies and standard deviations for SA1-DTC, FA-DTC and vibro-DTC models.

The dynamic model trained with SA-I data, SA1-DTC, was able to correctly predict texture in all trials. As with the SA1-STC-*slide*, this model was trained with samples collected at all 10 speeds. Meaning that, for each test trial, the exact stimulus condition shown to the model

had also been used to train the network. Our explanation for the exceptional performance of SAl-STC-*slide* under conditions of *all speed testing* (Section 5.4.2.2) also holds true for dynamic touch: Tactile data collected under identical stimulus conditions exhibits minimal variance and, therefore, spatial features which vary simultaneously with speed and texture can be learned and used for texture classification.

Again, mirroring the static case, FA-DTC performed comparatively worse than the equivalent SA-I model, SA1-DTC and similarly the explanation for this drop in performance is the relative inconsistency in FA tactile images collected at the same speed (see Fig. 5.5D).

The all speed, dynamic model trained with FA data, FA-DTC, improved upon its static counterpart, FA-STC-*slide* (Fig. 5.7B), suggesting that some degree of salient information is contained within the temporal patterns of artificial FA afferent firing. The overall prediction accuracy of FA-DTC was ~77% correct. The modal predicted class was always the true class. The variance in prediction, averaged across classes, was ~4 classes.

Finally, considering *all-speed* dynamic texture perception using our engineered vibrational channel, Fig. 5.11C. The trained model, vibro-DTC, was able to correctly predict texture in all but 1 trial. Clearly this suggests that using vibrational data collected via the TacTip's microphone is a viable approach to tex-

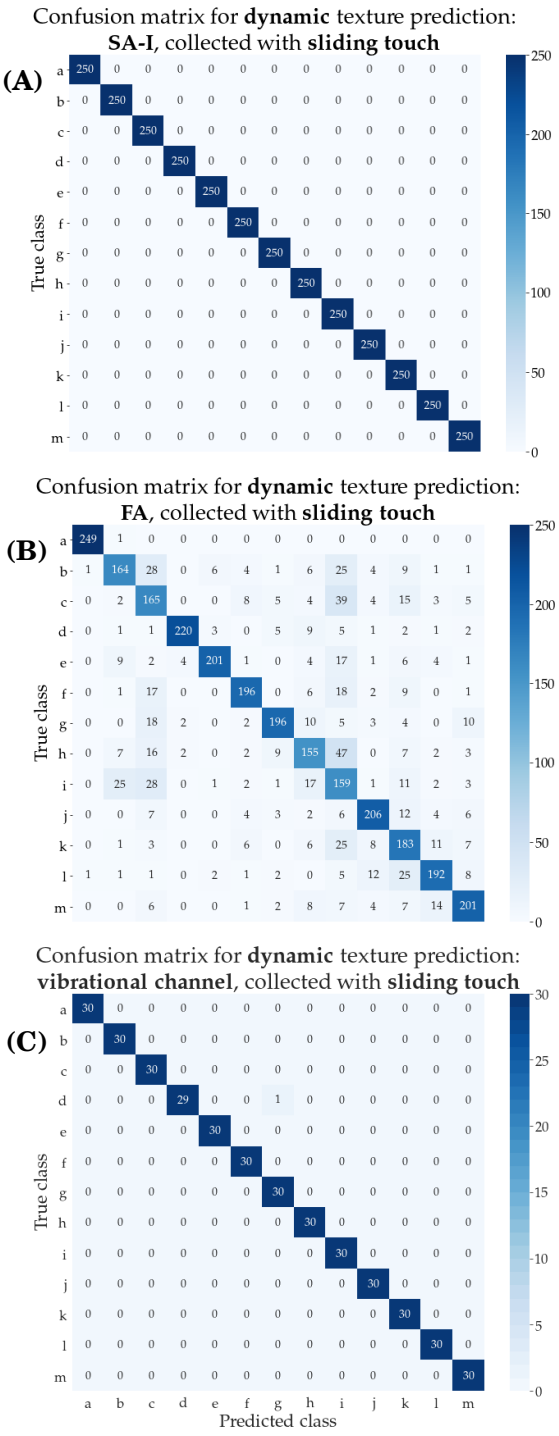

Figure 5.11: Confusion matrices of dynamic texture prediction for *all speed testing*. **A**, **B** and **C** show results for SA1-DTC, FA-DTC and vibro-DTC trained on SA-I, FA and vibrational channels respectively.

ture perception, at-least where the classifier is tested with data collected at speeds which were also used when training. Furthermore, the result suggests that the shape of frequency spectra of induced vibrations is a viable cue for texture. Therefore, our robotic system successfully uses temporal coding of vibrational cues for the classification of these textures.

As with SA-I models, it is likely that the high performance observed when using vibrational data under conditions of *all-speed-testing* is a result of the consistency in data across identical stimulation conditions; i.e., when both speed and texture align. In this case, the model will learn to use, not just the shape of the FFT samples, but also the absolute positions of spectral features, a property observed to vary with speed (see Fig. 5.10).

Texture Classification - *Leave-one-speed-out*

Prediction results for *leave-one-speed-out cross validation* models SA1-DTC-$\boldsymbol{v}$, FA-DTC-$\boldsymbol{v}$ and vibro-DTC-$\boldsymbol{v}$ (models described in Table. 5.2) are shown as confusion matrices in Figs. 5.12A, 5.12B and 5.12C respectively. Table 5.4 provides accuracies and standard deviations for SA1-DTC-$\boldsymbol{v}$, FA-DTC-$\boldsymbol{v}$ and vibro-DTC-$\boldsymbol{v}$ models averaged across hold-out speeds.

Firstly, considering SA1-DTC-$\boldsymbol{v}$ models, trained with SA-I data: Clearly performance was significantly worse than in the case of *all speed testing*, as was also the case for static texture perception. Similarly, we understand this to be a consequence of the variance in tactile images across speeds (Fig. 5.5B).

Each SA-I model performed at-least moderately better than chance. Accuracy, averaged across models with each hold-out speed, was ~50 % which matches that of the static variant (SA1-STC-*slide*). However, confusion matrices suggest a slight drop in performance when comparing the results of dynamic SA-I models to their static counterparts (Fig. 5.7A). This suggests that models based on SA-I afferent firing do not benefit from the addition of a temporal dimension; i.e., artificial SA-I afferents do not encode texture temporally. The small drop in performance is likely to be a consequence of a 10-fold reduction in the number of training samples, compared with static data. Fewer samples often results in over-fitting to the training data.

FA *leave-one-speed-out* models (Fig. 5.12B) clearly benefited from the availability of temporal features. The accuracy of the FA-DTC-$\boldsymbol{v}$ models was ~70 % compared to ~51 % for the static case (FA-STC-*slide*, Fig. 5.7). This suggests that artificial FA can encode texture in the spatio-temporal modulation of its firing; i.e., there is some benefit to extracting spatio-temporal over purely spatial features. This is in contrast with what was seen with SA-I artificial afferents, a result which was predicted from inspection of the tactile images (Fig. 5.5D vs. Fig. 5.5B), where more variation was seen within a single dynamic sample (i.e., a row) of FA firing compared to SA-I.

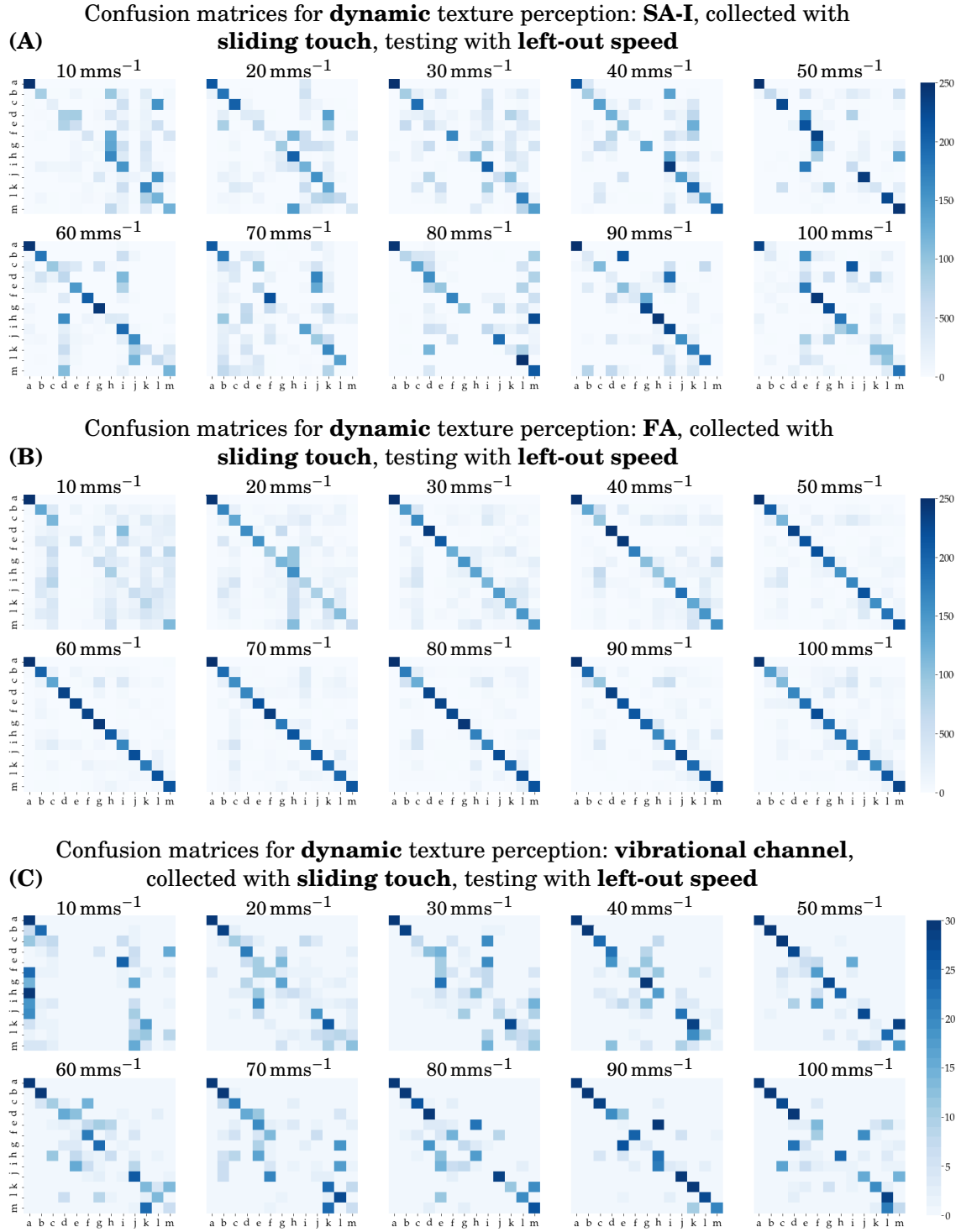

Figure 5.12: Confusion matrices of dynamic texture prediction using *leave-one-speed-out cross validation*. **A**, **B** and **C** are for SA1-DTC-$\boldsymbol{v}$, FA-DTC-$\boldsymbol{v}$ and vibro-DTC-$\boldsymbol{v}$ trained with SA-I, FA and vibrational channels respectively. Each matrix refers to a different model trained and tested with a held out speed, indicated by its individual title.

As with static perception, dynamic FA models (FA-DTC-$\boldsymbol{v}$) performed better above a threshold hold-out speed (50 mms^{-1}) than when using the *all speed testing* model (FA-DTC). The hypothesis presented for this with static perception was that low speeds are comparatively ineffective for texture perception, therefore, by including them in the test set, they contribute towards lowering the overall accuracy in the FA-DTC model.

We hypothesise that higher hold-out speeds produce better perception because each sample is formed of a fixed length time window (~0.1 s); when the stimulation speed is greater, more of the stimulus will pass the TacTip per sample. With more of the stimulus being felt, naturally each sample will contain more information.

Finally, we consider *leave-one-speed-out* using the vibrational channel (Fig. 5.12), vibro-DTC-$\boldsymbol{v}$. These models underperformed when compared to the vibro-DTC: average accuracy of ~52 % vs. 100 %. This suggests the developed method did not generalise well to unobserved speeds. Our hypothesis was that the 1D-CNNs would learn, not to use the exact shape of each vector, but rather a scale invariant representation of the structure, since our analysis of the data (Fig. 5.10) suggests that the structure is consistent within textures but sliding speed affects scaling. Although these models were capable of predicting texture in some cases, particularly, as predicted, for the smooth stimulus (class $\boldsymbol{a}$), the models clearly did not effectively extract speed-invariant features from FFTs.

5.5.2.1 Vibration Data Augmentation

To enable the vibration texture classification models to generalise over different speeds, clearly some measure must be taken to force the models to learn speed-invariant features. One approach to this would be to simply collect more data, specifically at a more complete set of speeds. This would likely reduce over-fitting by: *i)* increasing the amount of training data and *ii)* introducing more noise into the dataset. Furthermore, with a more continuous set of speeds, the held out speed is likely to look more similar to a portion of the training data simply because that data was collected at more similar speeds.

Alternatively, the existing data could be augmented which artificially produces new training data. Augmentation generally involves introducing variation into existing data by transforming it in some way; e.g., in image data, this could be by stretching, mirroring, rotating, zooming, translating or adding Gaussian noise [290]. Augmentation is desirable since it is considerably less time consuming than collecting more data; therefore, it is an approach we intend to implement.

Rather than randomly transforming existing data, however, we propose a more informed approach that is in effect very similar to collecting more data at a larger variety of speeds. Essentially, by 'stretching' and 'squashing' the existing FFT samples in the frequency domain, data collected at different speeds can be simulated.

Augmentation Procedure

The *leave-one-speed-out cross validation* method (Section 5.5.1.3) was used for constructing train, validation and test sets. Thus, we created 10 distinct sets with a different held out speed. Feature extraction was also performed as explained in Section 5.5.1.3.

For every 'real' training sample, six augmented samples were created, three of which were stretched in the frequency domain by a factor uniformly sampled between 1 and 2 and the remaining three were stretched by a factor uniformly sampled between 0.5 and 1 (effectively squashing). Therefore, we created $3120 = 6 \times N_{\text{train-real}}$ training samples per class, N_{train}, where $N_{\text{train-real}} = 520$ is the number of real training samples per class; i.e., the number of samples per class used to train vibro-DTC-$\boldsymbol{v}$.

Fig. 5.13 shows real and augmented vibration samples collected on stimulus $\boldsymbol{i}$. Augmented samples, derived from a real sample collected at 40 mms^{-1} (centre), are stretched by factors of 0.5 and 2, bottom left and right respectively. Notice how the augmented data has been scaled to appear similar to the real samples above, collected at 20 and 80 mms^{-1}, top left and right respectively. Through augmentation, samples collected 20 and 80 mms^{-1} have been artificial simulated.

Augmentation CNN Model

The same 1D-CNN model architecture was used as with real data, shown in Fig. 5.9. As with the *leave-one-speed-out cross validation* approach, 10 separate models of the same architecture were trained and tested using the procedure described in Section 5.5.1.3. These models are named vibro-DTC-$\boldsymbol{v}$-*aug*.

Results

Classification results of the models trained with augmented vibration data are shown as confusion matrices in Fig. 5.14. Table 5.4 provides accuracy and standard deviation for model and vibro-DTC-$\boldsymbol{v}$-*aug* averaged across hold-out speeds.

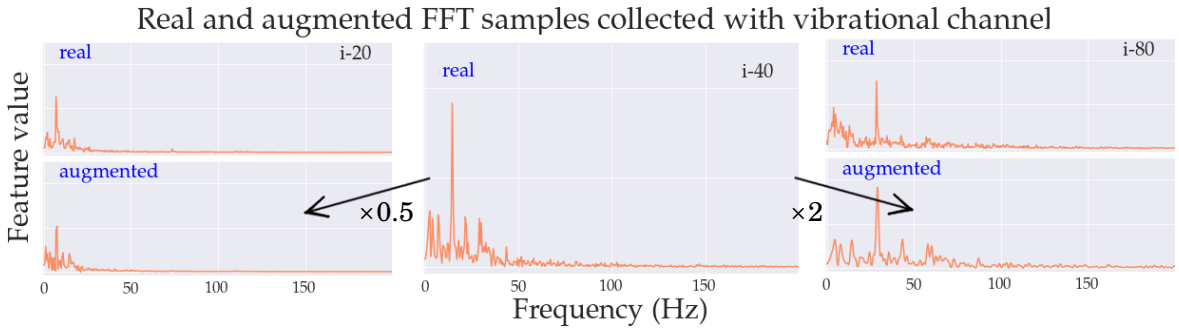

Figure 5.13: Real and augmented vibrational samples. All 5 plots show data collected on stimulus $\boldsymbol{i}$. *Centre*: real data, collected at 40 mms^{-1}. *Top left*: real data, collected at 20 mms^{-1}. *Top right*: real data, collected at 80 mms^{-1}. *Bottom left*: augmented data, stretched by a factor of 0.5. *Bottom right*: augmented data, stretched by a factor of 2.

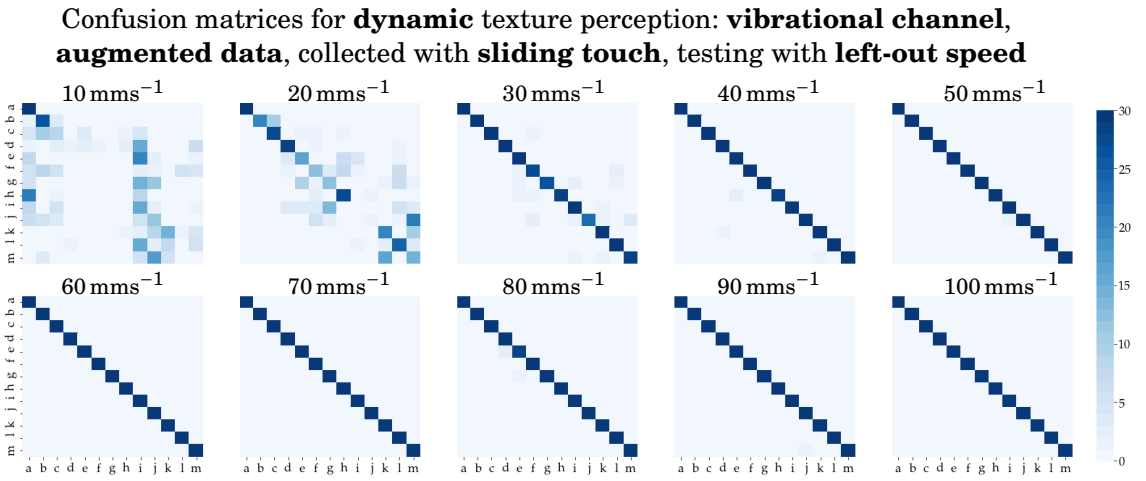

Figure 5.14: Confusion matrices of dynamic texture prediction using *leave-one-speed-out cross validation* with models trained on augmented vibration channel data: vibro-DTC-***v****-aug*. Each matrix refers to a different model trained and tested with a held out speed, indicated by its individual title.

Models tested with held out speeds 10 and 20 mms^{-1} show comparable performance to vibrational models trained with real data at the same held out speed (Fig. 5.12C), however, at all other speeds, there is a considerable improvement in accuracy over models trained with real data: accuracy is perfect, or close to, at hold-out speeds >30 mms^{-1}.

The improvement observed with augmented FFT samples provides strong evidence in support of the hypothesis that scanning speed in effect scales frequency spectra whilst retaining harmonic structure. In fact, the relatively poor performance at slower hold-out speeds can also be explained in terms of this concept: Consider the model tested on the hold-out set collected at 10 mms^{-1}, vibro-DTC-10-*aug*. This model was trained only on data collected at speeds >10 mms^{-1}. Therefore, according to the scaling hypothesis, training data that is similar to test data can only be produced by 'squashing' the FFT samples in the frequency domain. Squashed data only accounts for half of the training data since the other half was 'stretched', so it is comparatively unlikely for augmented data to look similar to test data. Furthermore, because the next highest speed is twice as fast, the only way to closely simulate a sample collected at >10 mms^{-1} is to squash a sample collected at 20 mms^{-1} by a scaling factor that is close to the maximum value of 2 (equivalent to stretching by a factor of 0.5). Since scaling factors are uniformly sampled, this further decreases the amount of training data that can accurately simulate test data. The described effect is true but to a lessening degree for increasing hold-out speeds.

5.5.3 Discussion

The availability of a temporal encoding of texture within artificial SA-I and FA afferents was examined by extending samples collected with sliding to include a temporal dimension and using

model architectures which can learn temporal as well as spatial features (ConvLSTM).

We found that artificial FA afferents provided a viable temporal code for texture, whereas artificial SA-I afferents did not. This is observed as an improvement in classification performance on held out speeds when using dynamic texture classification models, FA-DTC-$\boldsymbol{v}$, over static models, FA-STC-*slide*-$\boldsymbol{v}$.

Artificial FA afferents are more dynamically sensitive than artificial SA-I; e.g., the change in afferent firing between adjacent frames is likely to be more drastic, which is made apparent when comparing Figs. 5.5B and 5.5D. Indeed, it is understood that FA afferents are particularly responsive to vibrations (5-40 Hz) [49], which are commonly low amplitude compared with deformation associated with static touch. The heightened sensitivity to small scale dynamic stimulation may explain why artificial FA afferents were able to encode texture temporally.

The vibrational channel was tested for its ability to transduce and encode textural information in temporal patterns of recorded acoustic vibrations. Samples constructed as FFTs from ~2 s time segments (Fig. 5.10) of sliding tactile data were used as feature maps and temporal information was extracted using 1D-CNNs (Fig. 5.9). This method was found to be very effective when the model was tested with data collected at speeds which had also been used to train the network (Fig. 5.11C), however, was comparatively ineffective on held out speeds (Fig. 5.12C) because the model was unable to robustly learn speed-invariant features, identified broadly as the harmonic structure (scale invariant shape) of the FFTs.

Table 5.4: Table of results for all dynamic texture classification models. Variance here is the variance of prediction within a class, averaged accross all classes. Where the models are trained and tested with a *leave-one-speed-out cross validation* (indicated with $\boldsymbol{v}$), the results are averaged across all left-out speeds.

Model	Accuracy (%)	Variance (classes)
SA1-DTC	100	0.0
FA-DTC	77	4.0
vibro-DTC	100	0.0
SA1-DTC-$\boldsymbol{v}$	50	6.5
FA-DTC-$\boldsymbol{v}$	70	5.2
vibro-DTC-$\boldsymbol{v}$	52	3.7
vibro-DTC-$\boldsymbol{v}$-*aug*	88	1.3

We were able to drastically improve these models ability to generalise across speeds through data augmentation. Essentially, by stretching and squashing feature maps, different speeds are simulated (Fig. 5.13) which can be used to train the network, thereby enabling better classification of data collected at previously unobserved speeds. The improvement of texture classification through data augmentation in this way, appears to confirm that the harmonic structure may be leveraged as a feature for texture perception. This is an interesting result since it seemingly aligns with a compelling theory to explain speed-invariance of texture perception in humans: cortical computations are capable of extracting harmonic structure from neuron firing [131, 132], an idea that was inspired by observed timbre-invariance in auditory percep-

tion [133].

It is widely considered that the Pacinian system is the primary peripheral channel for mediating information about fine surface texture [128, 129, 291] and, in fact, it has been proposed that the high-frequency vibrations, transduced only via the PC channel, elicited by dynamic touch are necessary for discrimination of textures with an element size smaller than ~100 μm [57]. The stimuli used here may not be considered fine enough to produce a disparity in performance between artificial FA and the vibrational channels in dynamic texture classification. Therefore, we argue that with finer textures, the ability for the FA channel to be used for texture classification under dynamic conditions may be hindered, where, instead the vibrational channel might continue to provide viable temporal codes.

5.6 Concluding Remarks

In this chapter, we presented models for artificial texture discrimination based on theories of peripheral encoding of texture in humans. Here we assess the performance of artificial SA-I and FA channels for texture discrimination when statically stimulated and all three artificial tactile channels (SA-I, FA and vibrational) when dynamically stimulated.

We predicted that static perception models would perform poorly for fine textures, in agreement with duplex theory of human texture perception [57]. In particular, spatial response profiles (SRPs) (Section 4.6.3, Chapter 4) suggested that texture classification would be poor where bumps were spaced closer ~4 mm (classes ***a***-***i***). Whilst the static perception model trained with artificial FA pressing data (FA-STC-*press*) performed poorly over all textures, the equivalent SA-I model performed considerably better than expected on fine textures. As stated, this is likely because the network, SA-STC-*press*, was able to extract some feature or features from the tactile images which are not a direct consequence of the bump size or spacing in texture stimuli, but rather implicitly generated in the experimental setup. This is an unfortunate limitation of using techniques which are free to automatically learn features. Arguably, a more suitable approach may have been to use hand-crafted features based on spatial structure of tactile images, as certainly this would produce results more in line with the prediction that performance would significantly degrade where bump spacing was below ~4 mm. However, the availability, or otherwise, of spatial features resulting from bump size and spacing is obvious from the data and we feel that to use hand-crafted features based on this observable structure provides less contribution to the field than using modern techniques which are more in-line with the current state-of-the-art.

With the discussed limitation of the CNN method in-mind, we sought to make some tentative judgements about any trend in classification performance: On the coarsest textures (classes with bump spacing >~4 mm) accuracy was significantly better than that of the finest three textures (classes with bump spacing <~0.8 mm) and the average for the whole set. This suggests that as stimuli become rougher the added availability of spatial features improves accuracy above the

baseline level seen on the smoothest stimuli. This result seems to concur, in-part, with Katz's duplex theory [57, 125] of human texture perception, which states that the ability for humans to perceive fine textures with static touch is affected by the physical resolution of primary afferents in the human finger.

SA-I and FA models of static perception collected with sliding touch did not demonstrate any significant degradation of performance with reduced bump spacing. We believe this is because rather than spatial cues, the models were using frictional cues which are visible as shearing effects in the tactile images collected with sliding. In fact, studies have shown that human surface discrimination may employ a 'sticky-slippery' dimension which relies on frictional cues [292]. Regardless of the specific stimulus cues, the results of our static perception models suggest that information pertaining to texture is encoded within spatial modulation of artificial SA-I and FA afferents.

Leveraging an additional temporal dimension improved accuracy for all artificial FA models (FA-DTC/FA-DTC-$\boldsymbol{v}$) whereas it had little impact on artificial SA-I models. This suggests that artificial FA afferents may be more suited to temporal encoding than artificial SA-I. Neurophysiological experiments have shown that FA peripheral fibres respond more robustly than SA-I to dynamic stimulation with a range of natural textures. Furthermore the frequency composition of FA response reflected that of oscillations in skin induced by scanning [129], a finding which has informed many to believe FA afferents are leveraged for mediating information about texture through a temporal code. Therefore, the result that artificial FA afferents provide a spatio-temporal encoding of texture and artificial SA-I afferents do not, may be indicative of their natural counterparts.

Harmonic structure of data collected via the vibrational channel provided a viable cue for texture perception, although, the 1D-CNN was poor at learning speed-invariant features, demonstrated by the drop in performance when testing with held out speeds (vibro-DTC vs. vibro-DTC-$\boldsymbol{v}$). We identified that sliding speed did not affect the overall shape of FFTs, however, it did affect absolute frequencies; i.e., positions of the peaks. By augmenting data through 'stretching' and 'squashing' features maps, we were able to simulate data collected at extra speeds. This dramatically improved texture perception on held out speeds. Ultimately, the engineered vibrational channel was capable of accurately discriminating textures based on vibrational cues and we suggest, therefore, that it might offer a plausible model for the natural Pacinian system.

It may be argued that by simulating the held out speed through data augmentation, speed-invariance is achieved only in a superficial sense, since, in-fact, the model has effectively been trained on some samples which appear very similar to the held out speed. We concede that, if a chunk of the samples which simulate speeds closest to the held out speed were removed from the training data, performance would significantly drop and thus acknowledge that it is not the CNN model architecture which enables speed-invariance of texture perception. However, due to the continuous nature of the frequency domain scaling during the data augmentation

process, the model would have high accuracy if tested at any speed between 30 and, at least, 100 mms^{-1}. Furthermore, it could be argued that all parts of the training process, downstream of data collection (construction of FFT samples, data augmentation, CNN-feature learning and NN-decoding), are part of the perceptual system, regardless of whether the system extracts hand-crafted or automatically learned features. This perceptual system, considered in isolation, clearly enables speed-invariance. Whilst many theories suggest speed-invariance within human texture perception is achieved by using the approximate shape of frequency spectra of texture induced vibrations [131], precisely how these vibrational cues are encoded within the nervous system, and indeed, how harmonic structure is extracted is an open area of research [**?**]. Therefore, it is unreasonable to rule out the possibility of cortical processes that perform scaling of harmonic features analogous to the presented data augmentation technique. Thus, to understand the degree to which the presented method is truly biomimetic is an interesting avenue for future work but is also underpinned by our current understanding of biology.

We argue that the most appropriate way to further progress the state-of-the-art is to consider the types of tactile cue and encoding mechanism in detail, thus enabling systems which target these exact cues and therefore maximise the salience of this information and minimise redundancy.

In contrast to this approach, much of the previous work into texture classification has focused primarily on achieving high classification accuracy on real-world textures without such a rigorous consideration of the tactile cues enabling this [147, 149, 193, 219, 228–230]. This difference in perspective is exemplified by considering the recent work of Taunyazov et al. (2019) [229], where pressing and sliding were both used to stimulate the iCub's tactile forearm. The authors compared a number of different methods for classifying 23 natural textures including using hand crafted features, CNNs and LSTMs. The best performing method was to combine CNN-LSTMs trained on raw pressing and sliding data. It is hardly surprising that this approach performed the best when features are automatically learned from rich spatio-temporal data, however it is possible that a more bio-inspired targeted approach; e.g., harmonic structure of vibrotactile data, may have been more data efficient.

The GelSight has been used with great success for natural surface classification using statistical features encoded in the spatial structure within tactile images [193] and more recently CNNs to automatically learn features [203]. It is likely that humans would employ dynamic stimulation in order to discriminate the surfaces used in both studies due to their fine grained nature. The GelSight provides extremely high-fidelity tactile images which enables static texture perception on these natural surfaces. We argue that dexterous robots will be capable of dynamic stimulation and therefore be able to induce highly salient vibrotactile cues, thus the high-resolution of the GelSight becomes redundant.

The work of Fishel and Loeb (2012) [259] is aligned more closely with our philosophy. Numerical features were extracted from texture induced vibrations, measured using the BioTac's

hydro-acoustic pressure sensor, that have been proposed to encode subjective properties of texture: *roughness*, *fineness* and *stickyness*/*sliperyness* in humans [128]. This approach achieved impressive accuracy of 95% on 117 natural textures, which may be a consequence of the acute and highly salient features used: in-fact, the three subjective dimensions have been shown to be orthogonal in human perception [221]. Fishel and Loeb, however, did not consider spatial cues believed to be leveraged for texture discrimination of coarse textures. For future progression of the work presented in this chapter towards natural texture classification, a consideration of how these subjective properties may be encoded within tactile data, may be a beneficial approach as textures become more complex.

As mentioned in the literature review section (Section 2.2.2.6), to our knowledge, speed-invariance of texture perception has only been tackled in robotics in one other study; Romano and Kuchenbecker (2014) [230], where stimulation speed was applied as an additional feature to an SVM classifier. In contrast, our approach does not require any speed information: we identified a speed invariant feature of texture-induced vibrations (harmonic structure) and demonstrated that it could be used for highly accurate speed-invariant texture perception. This aligns with leading theories of speed-invariant texture perception in humans [131, 133]. We used 1D-CNNs to learn features within the shape of frequency spectra and decode texture class, which is a completely novel and potentially fruitful technique for robotic texture perception. Again, this demonstrates the value in highly informed feature engineering and particularly in using bio-inspiration.

By examining the data in this and the preceding chapter and cross-referencing with results from this chapter, we have identified that; *i)* spatial cues are available, when statically stimulating, down to the limit of the resolution of the sensing device and that these are encoded spatially in artificial SA-I afferents; *ii)* frictional cues are available when dynamically stimulating and these are encoded spatially in artificial SA-I and FA afferents; *iii)* temporal variation in the frictional cues provides information which is encoded spatio-temporally in artificial FA afferents; and *iv)* vibrational cues are available when dynamically stimulating and texture is encoded in the harmonic structure of these induced vibrations. Armed with these principles, it is conceivable how future methods can target these aspects of the data in order to improve tactile sensing in general. The natural progression of the work presented in this chapter is an application to natural textures and we believe that the methods developed within this chapter will provide a strong fundamental basis on which to develop an accurate and practical bio-inspired solution. Furthermore, we argue that, in this and the preceding chapter, we have demonstrated the possibility to engineer distinct artificial tactile channels which are plausible analogs of natural SA-I, FA and PC channels and that the concept of considering the availability of different tactile cues depending on stimulation type, stimulus scale and sensor technology and the associated encoding mechanisms will have benefits beyond texture classification, notably for the general application to robot manipulation.

CHAPTER 6

APPLICATION TO ROBOT HANDS

This chapter presents the development of a high-definition, biomimetic tactile fingertip, based on the TacTip, and its integration with a three-fingered, fully-actuated, robot hand: the Shadow Robot Company's Modular Grasper (Fig. 6.1).

We extract features from high-dimensional raw tactile images and infer information relevant to grasp quality using simple algorithms and relatively small amounts of training data in comparison to documented data-driven approaches. We also demonstrate the ability for these percepts to be leveraged in an online grasp adjustment procedure.

The work presented in this chapter has been published in the following peer-reviewed publication (a contribution statement is included):

- Pestell, Nicholas, Luke Cramphorn, Fotios Papodopolous and Nathan F. Lepora. "A sense of touch for the Shadow Modular Grasper." IEEE Robotics and Automation Letters 4.2 (2019): 2220-2226.

  Contribution statement: N. Pestell realised the concept and wrote the manuscript with advice from L. Cramphorn and N. F. Lepora. L. Cramphorn developed the tactile feature extraction technique. N. Pestell designed the sensor, developed the perception software and integrated it with existing control code and performed experiments. F. Papodopolous aided with system integration and experimentation.

6.1 Background

Robot hands have seen accelerated development in recent years [7], advancing attributes such as dexterity, grip strength and ease of use. Two-fingered grippers are deployed in large numbers for repetitive manufacturing tasks, whereas, more advanced, multi-fingered hands are yet to find

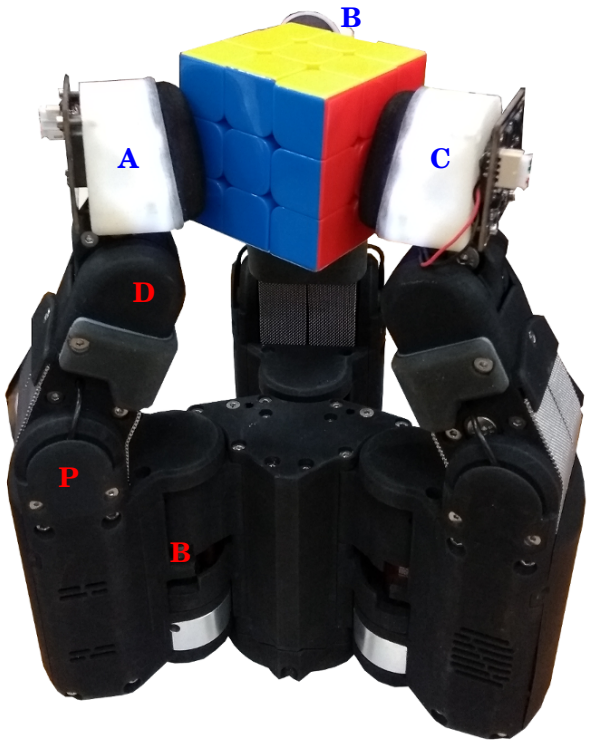

Figure 6.1: Image of the developed tactile sensors integrated with the Shadow Modular Grasper. Base, proximal and distal joints are labelled in red, **B**, **P** and **D** respectively. Tactile fingertips **A**, **B** and **C** are labelled in blue.

applications outside of research. Thus, a gap persists for automation of small scale production, where robots are required to grasp and manipulate unknown objects [1]. This gap can only be filled by dexterous, multi-fingered robot hands.

Given the advances in the state-of-the-art of robot hands, it is surprising such hands have not yet found widespread application. One contributing factor may be a lack of sufficient tactile sensing capabilities. Indeed, it is known that humans rely heavily on their sense of touch to maintain a stable grasp [293].

Here we provide a concise review of tactile sensing for robot hands, for the purpose of grasp stabilisation, in the context of the presented work. For a more in-depth review of tactile sensing with robot hands in general, the reader is referred to Section 2.2.3 of the literature review.

In 2011, Romano et al. [262] conducted an innovative study into a grasp control framework using touch as an integral component, using comparatively rudimentary hardware: two 5x3 capacitive tactile sensors and an accelerometer integrated with the PR2 two-fingered gripper. Different states within the control system were triggered by hard-coded tactile signals.

Since Romano's influential paper, there have been many attempts at improving grasp stability with the introduction of tactile sensing, primarily with data-driven approaches [192, 194, 194, 206, 294–296], these methods are often impractical due to the large quantities of training data required and poor generalisability. In general, the tactile sensors are low resolution, array-based technologies, which do not allow for extraction of high-level features with a more direct relationship to the object being held.

The bias towards these data-driven methods is partially due to low resolution of the tactile sensors employed. Without sufficient acuity, sensors provide un-intuitive representations of tactile contact. In contrast, optical tactile sensors; e.g., the GelSight [193] and the TacTip [13],

provide high-resolution tactile images which ease interpretation.

The TacTip has shown potential for integration with robot hands; e.g., [25] where a TacTip based tactile thumb (TacThumb) was integrated with the Open-Hand M2 gripper and [190] where two TacTip devices were mounted as fingertips on the Open-Hand GR2 gripper. In both of these studies, the authors utilise a supervised-learning method to achieve precise in-hand manipulation of custom-made objects.

Whilst the aforementioned literature shows clear benefits of tactile sensing for grasping, a common shortcoming is the nature of the tactile data available and/or the amount of data required to interpret it. Here we present a more flexible platform by integrating a highly sensitive, high-resolution, optical tactile sensor with a fully-actuated industrial robot hand. We demonstrate the potential for this system to improve grasps on unknown, real-world objects by presenting a method for grasp adjustment that could feasibly form part of a higher-level control framework.

6.2 Development of a Tactile Modular Grasper

6.2.1 The Shadow Modular Grasper

The Shadow Modular Grasper is fully actuated with 9 degrees of freedom (three per finger). The system is fully integrated with ROS and the user can control both specific joints and whole-hand grasps. Each identical finger has base, proximal and distal joints with one brushed DC motor per joint and can be easily attached and detached resulting in a modular system. Low level control

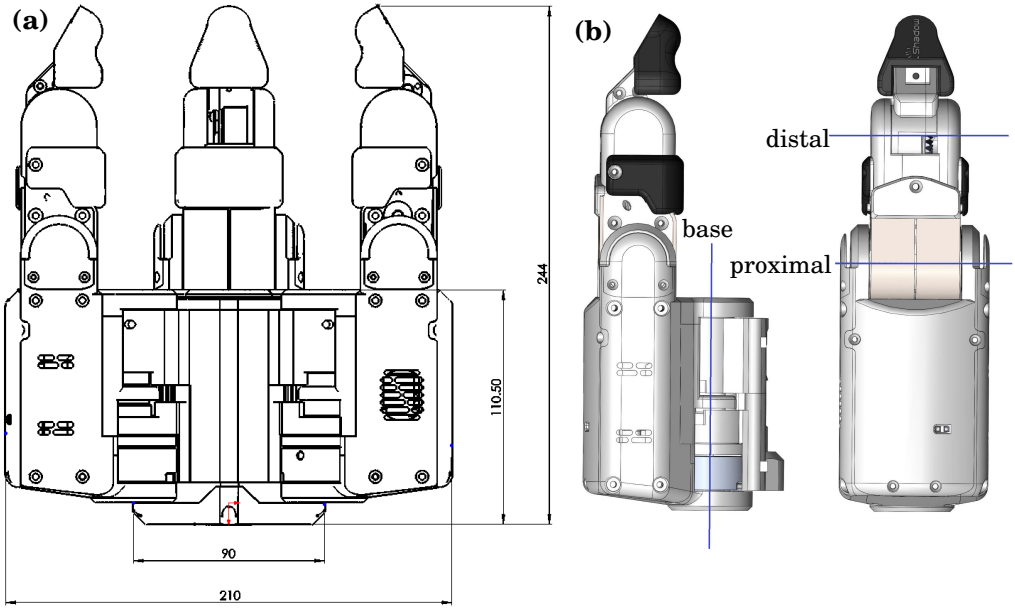

Figure 6.2: **(a)**: Drawing of whole Shadow Modular Grasper, showing dimensions in mm. **(b)**: Model of a single finger with base, proximal and distal joints labelled. Both figures taken from [27].

is performed with PWM on the voltage delivered to each motor. The full hand has a total mass of 2.7 kg and a payload of 2 kg and each finger can apply 10 N of normal force. Each joint has a dedicated torque sensor and encoder for closed loop control in torque and position mode. The grasper features a back-drivable gearbox enabling inherent compliance, which is an essential component when working in unstructured environments. The whole unit requires only two connections: power (24 V) and comms (EtherCAT) [27].

6.2.2 Tactile Fingertip Development

6.2.2.1 Hardware Design

Tactile sensing is enabled by replacing the fingertips of the Modular Grasper with three custom-built tactile sensors (Fig. 6.1). The fingertips are comprised of two main components: a compliant finger-pad which deforms when contacted and a rigid body which fixes the camera in place and houses other electronics (Fig. 6.3).

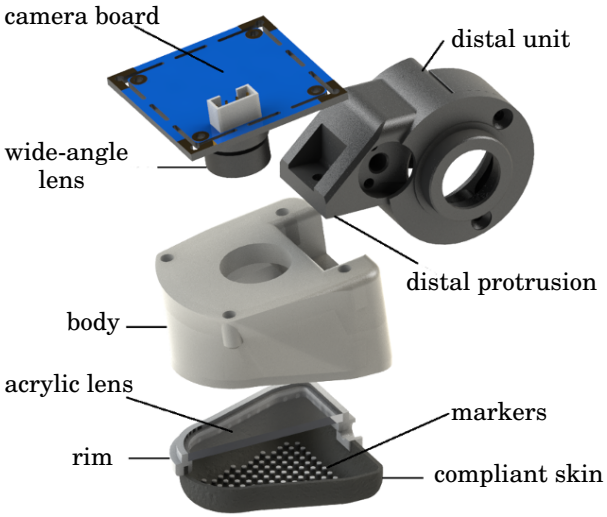

Figure 6.3: Exploded CAD model of the tactile fingertip.

The finger-pad is fully 3D-printed as a single part with a multi-material 3D-printer (Stratasys Objet). A sensing surface 'skin' is printed in Tango Black+ (Shore A 26-28). The inside of the skin is tessellated with a triangular pattern of 97 pins (Tango Black+, 3 mm length and 2 mm diameter). White markers on the end of pins are printed in rigid Vero White and provide a visual representation of the tactile stimulation. A rim (Vero White) enables a press-fit connection to the fingertip body. A clear acrylic sheet (2 mm thick) is glued into the rim resulting in a small cavity between the skin and the acrylic lens. This cavity is filled with a two-part cure, clear, silicone gel (RTV27905, Techsil UK (~Shore OO 10)) using a manual injection method. The gel helps to reduce hysteresis while still enabling deformation. The finger-pad shape closely matches the shape of the Modular Grasper's original fingertip contact surface.

The finger-pad is press-fit into a hollow body which is 3D-printed in ABS. A 2.0 megapixel CMOS array USB web-cam (ELP cameras) is mounted on the back via four M2 screws. The camera is used in HD mode (1920x1080). Dimensions of the fingertip body are optimized to enable a view of all markers whilst minimizing the overall depth of the fingertip. The markers are illuminated by four LEDs arranged on two PCB strips of two LEDs each. The PCBs are glued to the inside of the body, close to the interface with the finger-pad.

6.2.2.2 Hardware Integration

The aim is for modular tactile fingertips without modifying other parts of the hand. This introduces a set of major design challenges. Shown in Fig. 6.4, the distal unit fixes to the fingertip by way of a ~15 mm protrusion dictated by the original hand design. To overcome occlusion of the markers by the distal protrusion, the camera is shifted forward and fixed off-parallel from the sensing surface by an angle of ~8°. This geometry is unique from all previous versions of the TacTip, where the camera is mounted directly above and parallel to markers.

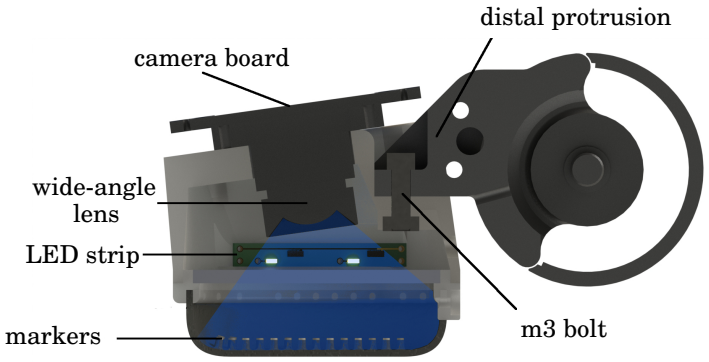

Figure 6.4: Computer modelled cross-sectional view of fingertip assembly.

Integration of three sensors has, to date, not been attempted with TacTip-based sensing. A solution proposed here is to connect each tactile fingertip to its own dedicated USB-hub. With three dedicated hubs, the data transfer occurs in parallel without reducing the frame rates, ~20 fps per camera.

6.2.2.3 Tactile Feature Extraction

Feature extraction is performed using a Voronoi method previously demonstrated to achieve inference of pressure and contact locations with the TacTip [247]. Marker positions are tracked with a simple blob detection algorithm implemented with OpenCV in Python [274]. A Voronoi tessellation is created over the sensor skin, treating the marker positions as seeds (Python, SciPy; scipy.spatial.Voronoi [297]) (Figs. 6.5a and b). The areas of each Voronoi cell are related to local skin deformation, where increased size corresponds to indentation which can be integrated to achieve volumetric change and can be mapped to pressure using simple 1D regression.

Visual representations of the surface deformation are obtained by interpolating the change in Voronoi cell areas over the fingertip. A centre-of-pressure, a tactile analogue of centre-of-mass, is computed as an average of marker positions weighted by their corresponding cell area (Fig. 6.5c).

6.3 Offline Perception

Orientation of the fingertip relative to the contact surface maybe of importance when grasping an object. For example, one might assume that grasp quality is compromised if the direction of the applied force by a individual finger is significantly offset from the normal to the surface of the

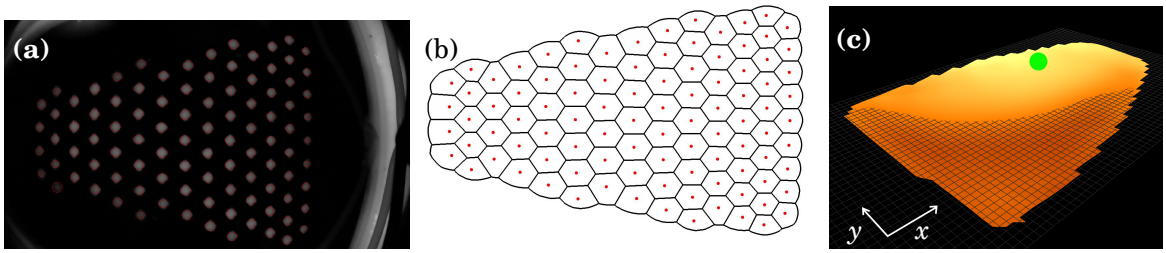

Figure 6.5: **(a)**: View of markers imaged by the camera and tracked with OpenCV. **(b)**: Voronoi tessellation over markers. **(c)**: Visual representation of surface deformations with centre-of-pressure shown as a green spot.

grasped object. Thus, here we examine the ability of the presented sensor and described feature extraction technique for perceiving roll and pitch relative to a flat surface.

6.3.1 Data Collection

The fingertip is mounted as an end-effector on a six degree-of-freedom robot arm (UR5, Universal Robotics). The sensor maintains continual contact with a flat acrylic plate and the robot re-orients the sensor relative to the plate. Data is sampled randomly from a 2D grid of roll, ϕ, and pitch, θ, values, $-16° \leq \phi \leq 16°$ and $-11° \leq \theta \leq 3°$.

Experimental set-up is shown in Fig. 6.6. ϕ and θ angles are equally space by 2° and 1° respectively yielding, a total of $N_\phi N_\theta = 17 \times 15 = 255$ data points.

Three seconds of training data (~ 60 frames) are collected for each sample. A sample consists of a time series of centre-of-pressure vales r_{ki}, where $1 \leq k \leq N_{\text{frames}}$ and $1 \leq i \leq N_{\text{dims}}$; $N_{\text{frames}} \simeq 60$ and $N_{\text{dims}} = 2$ for x and y positions.

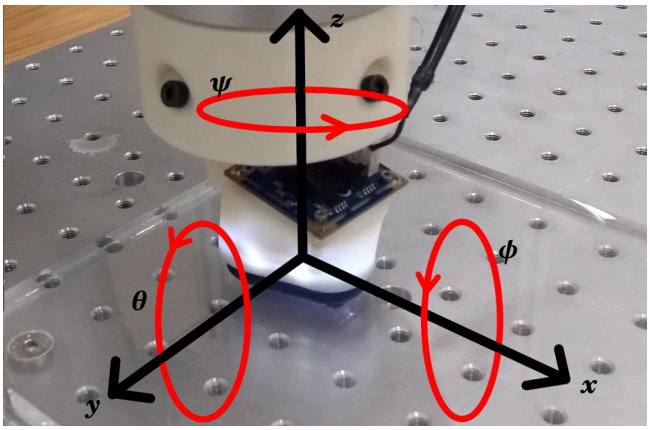

Figure 6.6: Data collection set-up with tactile fingertip mounted as an end-effector on a UR5 robot arm. Showing roll, ϕ, pitch, θ and yaw, ψ orientations relative to the sensor.

Three separate test sets are collected, on the same acrylic plate, of 200 data points each, sampled at random from a continuous space within the same range of ϕ and θ. Each set is collected at a different depth since we are interested in the degree to which our methods are pressure invariant. The depths, $-0.5\,\text{mm} \leq z_l \leq 0.5\,\text{mm}$, are equally spaced, where $z_l = 0\,\text{mm}$ corresponds to the depth used for training.

The ranges of ϕ and θ and of z are chosen in combination, to avoid damage to the sensor but also to provide examples of extremely light touch: pressure on the sensor is maximised when ϕ and θ are at extremes and z is minimised. The location of ϕ and θ ranges additionally provide a contact area which is maximised when both ϕ and $\theta = 0$.

To further examine the generalisability of our method we perform additional tests with data collected on two generic 3D-printed shapes: a dome and an edge. On the edge, we collect two sets with different yaw, ψ, angles: 0 and 90° respectively (Fig. 6.7). Data is collected using the same random sampling procedure as with the acrylic plate.

We repeat the training and testing data collection procedures once for each fingertip. This is because of small inconsistencies in the manufacturing procedure which may lead to different physical and optical responses.

6.3.2 Perception

Prior to training, data is averaged across frames, so each sample has $N_{\text{dims}} = 2$ features. We map centre-of-pressure-xy position to predictions of ϕ and θ via three separate multivariate, linear models: A simple linear model (1$^{\text{st}}$-order polynomial), 2$^{\text{nd}}$- and 3$^{\text{rd}}$-order polynomials.

Models are fitted using Python, Scikit-Learn (sklearn.linear_model.LinearRegression) [298]. Practically, Scikit-Learn uses Ordinary Least Squares to fit these models; Scikit-Learn wraps a Scipy implementation of Ordinary Least Squares (scipy.linalg.lstsq).

6.3.3 Results

We compared three competing model types (simple, 2$^{\text{nd}}$- and 3$^{\text{rd}}$-order linear models) for predicting ϕ and θ. Our model should accurately approximate training data whilst generalising to unobserved samples. Table 6.1 shows R^2 scores for three competing models for all three tips, on test data collected on the acrylic plate at each test depth (see Fig. 6.6), the dome and on the edge at $\psi = 0$ and 90° (see Fig. 6.7).

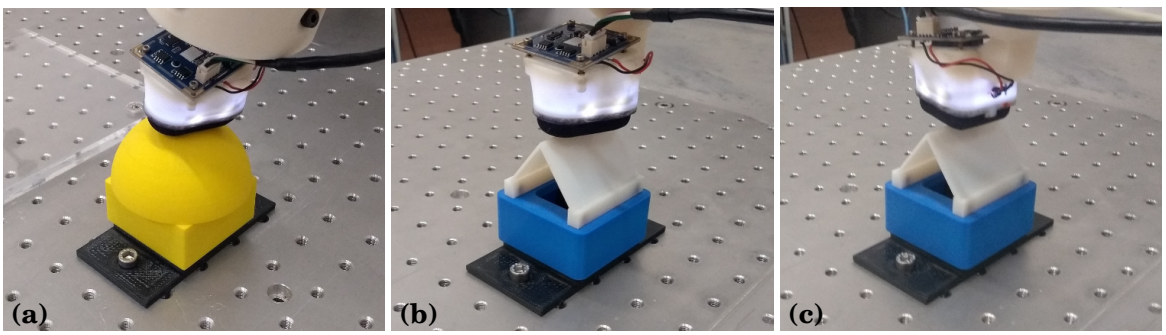

Figure 6.7: **(a)**: Data being collected on the dome stimulus. **(b)** and **(c)**: Data being collected on the edge stimulus at $\psi = 0$ and 90° respectively.

On the flat plate, where test sets were collected at training depth (0 mm), the 3rd-order polynomial model achieved the highest R^2 scores over the whole set (0.95, 0.94, and 0.93). This model, however, under-performed at the two unseen depths, particularly at 0.5 mm where it achieved the lowest three R^2 scores for the flat stimulus (0.67, 0.28 and 0.40).

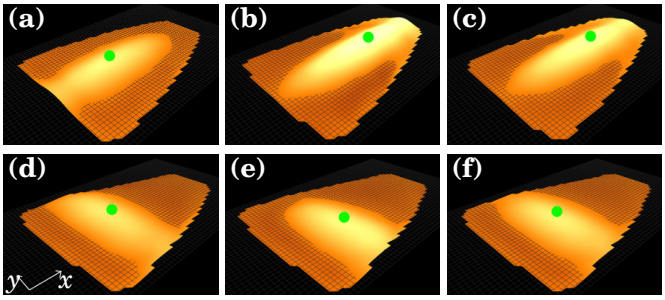

Figure 6.8: **(a)**, **(b)** and **(c)**: Visualisations of surface deformation with different values of ϕ and θ with ψ = 0°. **(d)**, **(e)** and **(f)**: Visualisations of surface deformation with different values of ϕ and θ with ψ = 90°.

The 3rd order model is expected to perform well on data close to the training set, being the most flexible. However, this flexibility is likely to be the cause of poor performance seen at other depths, as the model responds to details which are specific to the training data.

In contrast, both 1st- and 2nd-order polynomials performed more consistently on the flat plate: The lowest R^2 scores were 0.72 and 0.71 for the 1st- and 2nd-order polynomials respectively. This improved consistency is likely due to a lower variance in the models, giving better generalisation.

All three models are capable of generalising to the dome stimulus: all R^2 scores are above 0.6. As with the unobserved depths, the worst performing model was the 3rd order polynomial scoring 0.75 averaged across the three tips, vs. 0.88 and 0.86 for 1st- and 2nd-order polynomials respectively.

All three models, however, performed poorly on the edge at both ψ angle orientations: the highest R^2 score was 0.46. We believe this is caused by a singularity when exposed to edge like stimuli: the centre-of-pressure remains fixed in the dimension perpendicular to the edge regardless of ϕ and θ. Fig. 6.8 demonstrates this effect. The centre-of-pressure remains consistent in the x dimension when ψ = 0° and, contrastingly, it remains consistent in the y dimension when ψ = 90° regardless of ϕ and θ in both cases.

To test this hypothesis we examined R^2 scores attained when making separate predictions

Table 6.1: R^2 scores of ϕ,θ prediction for 1st-, 2nd- and 3rd-order polynomial linear regression for tips A, B and C, as labelled in Fig. 6.1, calculated for test sets on the flat acrylic plate at depths of -0.5, 0 and 0.5 mm, the dome and the edge at ψ = 0 and 90°.

	1st order						**2nd order**						**3rd order**					
Depth	**Flat plate**			**Dome**	**Edge**		**Flat plate**			**Dome**	**Edge**		**Flat plate**			**Dome**	**Edge**	
(mm)	**-0.5**	**0**	**0.5**	**0**	**0**		**-0.5**	**0**	**0.5**	**0**	**0**		**-0.5**	**0**	**0.5**	**0**	**0**	
ψ (°)	**0**	**0**	**0**	**0**	**0**	**90**	**0**	**0**	**0**	**0**	**0**	**90**	**0**	**0**	**0**	**0**	**0**	**90**
tip-A	0.91	0.90	0.91	0.92	0.42	-0.01	0.92	0.91	0.92	0.90	0.46	-0.07	0.84	0.95	0.67	0.82	0.28	-0.03
tip-B	0.87	0.89	0.72	0.87	0.42	0.25	0.89	0.92	0.71	0.87	0.46	0.29	0.85	0.94	0.28	0.83	0.32	0.29
tip-C	0.88	0.88	0.81	0.84	0.36	0.17	0.90	0.90	0.82	0.80	0.36	0.21	0.76	0.93	0.40	0.61	0.21	0.25

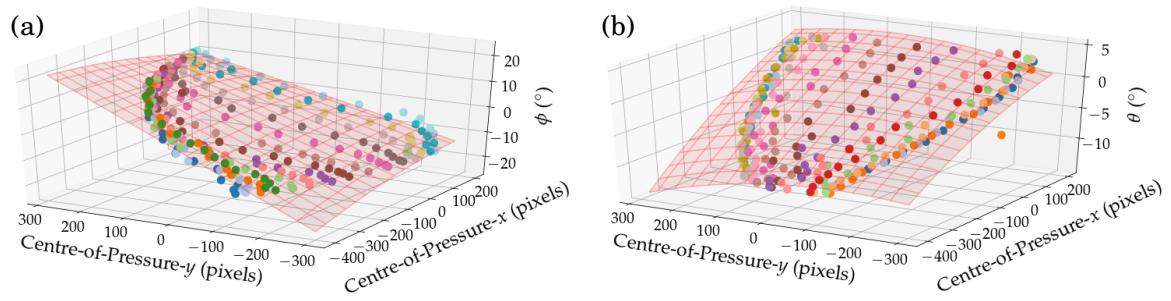

Figure 6.9: **(a)**: Scatter plot of xy-centre-of-pressure vs ϕ. **(b)**: Scatter plot of xy-centre-of-pressure vs θ angle. In both plots, the surface shows a 2^{nd} degree polynomial fit and each colour represents a constant θ and ϕ in **(a)** and **(b)** respectively.

of ϕ and θ for data collected on the edge. Results of this test are shown in Table 6.2. The three models perform well at $\psi = 0°$ when predicting θ but poorly when predicting ϕ. At $\psi = 90°$, good performance is achieved for ϕ predictions whereas the models perform poorly when predicting θ. In both cases, $\psi = 0°$ and 90°, better performance is observed when predicting the dimension which is parallel to the edge orientation, confirming the singularity hypothesis. We have, therefore, identified a limitation with sensing under these conditions.

Based on results from this data, we consider the 2^{nd}-order model to be the most suitable for predicting ϕ and θ with the application to robot hands.

Table 6.2: R^2 scores for separate predictions of ϕ and θ for 1^{st}- (simple), 2^{nd}- and 3^{rd}-order polynomial linear regression for tips A, B and C, as labelled in Fig. 6.1, calculated for test sets collected on the edge stimulus at ψ of 0 and 90°.

	1^{st} order				2^{nd} order				3^{rd} order			
ψ (°)	0		90		0		90		0		90	
Target	ϕ	θ	ϕ	θ	ϕ	θ	ϕ	θ	ϕ	θ	ϕ	θ
tip-A	0.33	0.89	0.88	-3.7	0.37	0.92	0.80	-3.7	0.18	0.78	0.76	-3.3
tip-B	0.32	0.87	0.84	-2.5	0.35	0.91	0.86	-2.4	0.23	0.77	0.89	-2.5
tip-C	0.25	0.91	0.96	-4.2	0.25	0.91	0.84	-3.2	0.11	0.70	0.92	-3.4

Visualisations of the 2^{nd}-order polynomial model for tip-B are shown in Fig. 6.9. The data is well ordered and the model appears a suitable fit. We observe strong correlations of centre-of-pressure-x and -y positions with θ and ϕ respectively. This is expected since the x- and y-axis align with θ and ϕ respectively. We also observe some correlation of centre-of-pressure-y and -x positions with and θ and ϕ respectively, suggesting both features are useful predictors.

6.4 Online Grasp Adjustment

Here we investigate the capabilities of the integrated Shadow Modular Grasper with tactile fingertips. We look at the capacity for perceived roll, ϕ', and pitch, θ' to be used in a grasp adjustment procedure (Section 6.4.1) and the potential for this to improve grasp quality on three objects from the YCB object set: a Rubik's cube, Pringles can and mustard bottle [299]. In this

context, ϕ and θ relate to fingertip roll and pitch angles with respect the surface, at the point of contact, of a grasped object.

6.4.1 System Integration

To integrate tactile feedback, we build upon a pre-existing grasp control system for the Shadow Modular Grasper.

The existing grasp controller implements a high-level strategy consisting of independent control of each joint through its dedicated motor (full actuation). Each joint can be controlled in either position or torque mode. The grasp controller implements a 1 kHz update loop which iteratively updates joint modes and joint targets via closed loop PID control.

Per finger, the grasp controller issues 3 torque demands, via EtherCAT, to a dedicated finger control board on the hand. A 5 kHz control loop running on the control board for each finger outputs 3 PWM voltage signals for respective base, proximal and distal joints. Encoders measure joint positions and torque which are sent back to the grasp controller via EtherCAT and used to compute error values for the PID.

We have written a python driver, which runs as three instances, one for each sensor. The drivers run on the host PC and interact with the grasp controller (C++), also running on the host machine, via a ROS-network. The sensor drivers perform acquisition of sensor data, feature extraction, contact detection and attain predictions, ϕ' and θ'. The high-level system structure is depicted in Fig. 6.10.

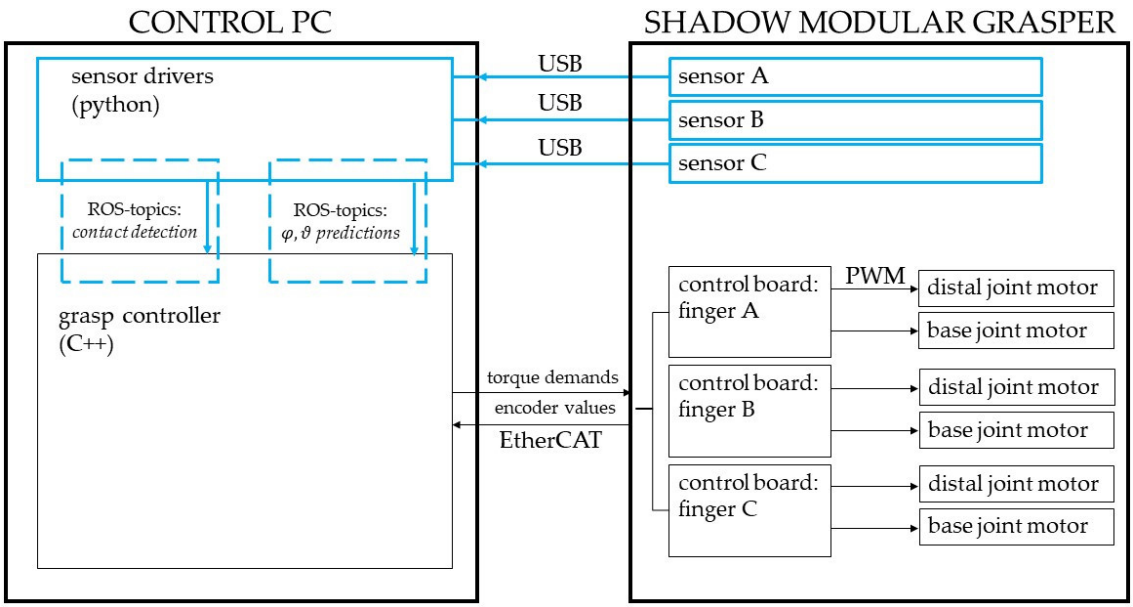

Figure 6.10: Diagram of the system structure. Novel contributions to the existing framework are highlighted in blue.

The generic grasp is adapted to be now include two distinct phases: *(i)* **closing phase** and *(ii)* **adjustment phase**.

(i) During the **closing phase** all joints are controlled in position mode and commanded to a set of target angles via a PID which is an implementation of the roscontrol ROS-package. The trajectory of each finger forms a symmetric grasp where net momentum is zero: base joints are fixed at the centre of their ranges (see Fig. 6.12a), proximal and distal joints slowly close (see Figs. 6.1 and 6.2 for reference of joint names).

Throughout this phase, the hand controller is listening to a set of 'contact-detection' ROS-topics, published by each sensor driver. Contact is determined by upward threshold crossing on the surface deformation (Fig. 6.5). Upon contact-detection each respective finger stops moving.

(ii) After all sensors have detected contact the **adjustment phase** is entered: the controller switches all proximal joints to torque mode and applies a fixed squeezing torque to the object. Base and distal joints remain in position mode and are servoed with a novel adaptation of the PID controller used in *(i)*. PID process variables are ϕ' and θ' for base and distal joints respectively, and their respective setpoints (targets) are ϕ and $\theta = 0$. A diagram of this closed loop control system is shown in Fig. 6.11.

An important concept is the mapping from sensor frames, in which ϕ and θ are defined, to joint-space, in which low-level control is performed: we treat this as one-to-one, where ϕ' and θ' are related directly to base and distal joint targets respectively. This concept is depicted in Fig. 6.12, where dashed blue lines and solid blue lines show joint positions pre and post adjustment respectively.

To steady the grasp, this adjustment phase has a time-out period, the length of which is obtained from experimentation (Section 6.4.2).

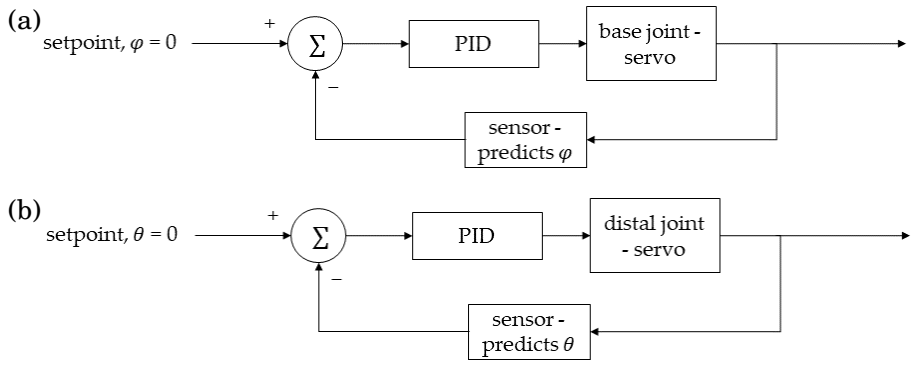

Figure 6.11: **(a)**: Diagram of separate closed loop PID controllers in phase *(ii)* (adjustment phase) for base and distal joints, **(a)** and **(b)** respectively. This system is duplicated three times: once for each finger.

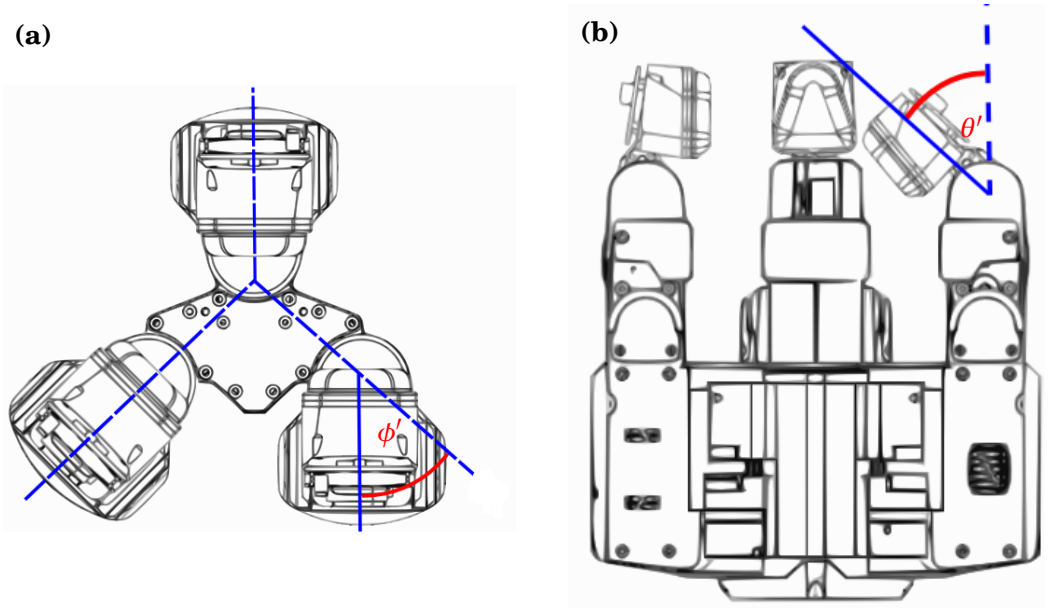

Figure 6.12: Technical drawings of the Shadow Modular Grasper equipped with tactile fingertips, showing how tactile percepts relate to joint angles and process variables in the control system. **(a)**: Plan view showing how ϕ' is related to base joint rotation. **(b)**: Side-on view showing how θ' is related to distal joint rotation.

6.4.2 Results

Informed by results in Section 6.3.3, ϕ and θ are predicted using a 2^{nd}-order polynomial regression model.

We have provided a supplementary video, available on YouTube [300] containing successful grasp attempts on all three objects.

Roughly half the trials were able to maintain a stable grasp throughout the entire adjustment phase. We will first analyse the behaviour for three typical successful grasps before evaluating the cause of failure in the remaining portion of our trials.

Fig. 6.13 shows successful grasps of all three objects. Alongside each image are tactile visualisations from each fingertip. The objects are initially held in place by a human participant before passing over to the robot when all three fingers have made contact. The top row shows images at initial contact detection (prior to tactile adjustment) and the bottom row shows images after tactile adjustment, ~10 s later.

The top and bottom rows of Fig. 6.13 show noticeable differences in both the grasp images and the tactile visualisations. In general, the grasp images show that the fingertips rotated around each object to minimise ϕ and θ. Inspection of the tactile visualisations suggests that overall deformation of each fingertip increased subsequent to adjustments in all cases. This suggests that, for these examples, the grasp controller performed as designed: to increase contact surface area and at each fingertip.

The hand made a good initial grasp of the Pringles can, owing to its symmetrical shape, with

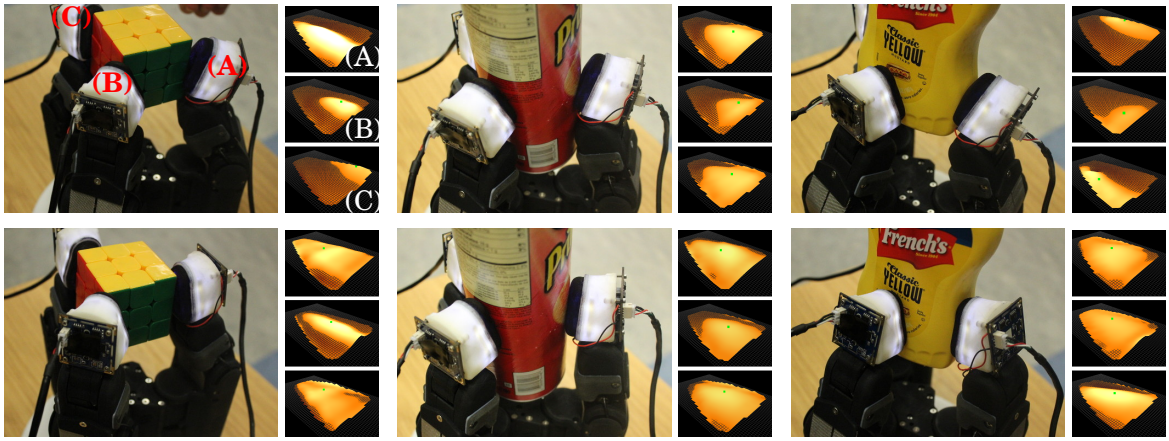

Figure 6.13: Images of the grasps on the Rubik's cube, Pringles can and mustard bottle, before and after tactile adjustment; top and bottom rows respectively. Tactile visualisations for the three fingertips are displayed to the right of each grasp image. Fingertips are labelled on the top left image and visualisations for reference.

ϕ close to zero prior to adjustment for all three fingertips. We observe a slight modification of θ for each fingertip which in-turn increased the contact surface area. The mustard bottle is the most irregular of the three presented objects. Despite this, the hand maintained a stable grasp throughout the adjustment phase and, increased the contact surface area on all three fingertips.

On the Rubik's cube, all three fingertips contacted edges resulting in singularities for predicting ϕ (Section 6.3.3). However, all three fingers moved to settle the centre-of-pressure in the middle of each finger-pad. For fingertips A and C we believe this is because initially the centre-of-pressure was towards the edge of the pad: despite this producing a measurement singularity, our algorithm perceives an off-set in ϕ, so the base joint is servoed to counteract the perceived off-set. This rolls the sensor onto the face of the Rubik's cube. Contrastingly, fingertip B maintained its location on the edge of the cube which we consider to again be a consequence of the perceived ϕ. The initial centre-of-pressure was close to the middle of the finger-pad, so our algorithm perceived ϕ to be close to zero. Whilst this grasp may not maximise the overall contact surface area, it is a stable configuration given its starting point.

Fig. 6.14 shows base and distal joint angles and centre-of-pressure-x and -y positions vs. time when successfully grasping the Rubik's cube, for fingers A, B and C, (as labelled in Fig. 6.13). All three fingers detect contact at roughly the same time. Subsequently, joint angles are adjusted before reaching the time-out, observed as a flattening of the blue curves. A time-out of 10 s appears to give the hand suitable time to re-adjust. During the same period, the centre-of-pressure migrates towards 0 in both dimensions. The shapes of these curves suggest that hand control performed as designed: to servo base and distal joints in order shifting the centre-of-pressure to the middle of each finger-pad. These plots demonstrate an interesting emergent behaviour: the base and distal joint angle curves mirror the centre-of-pressure-y and x positions

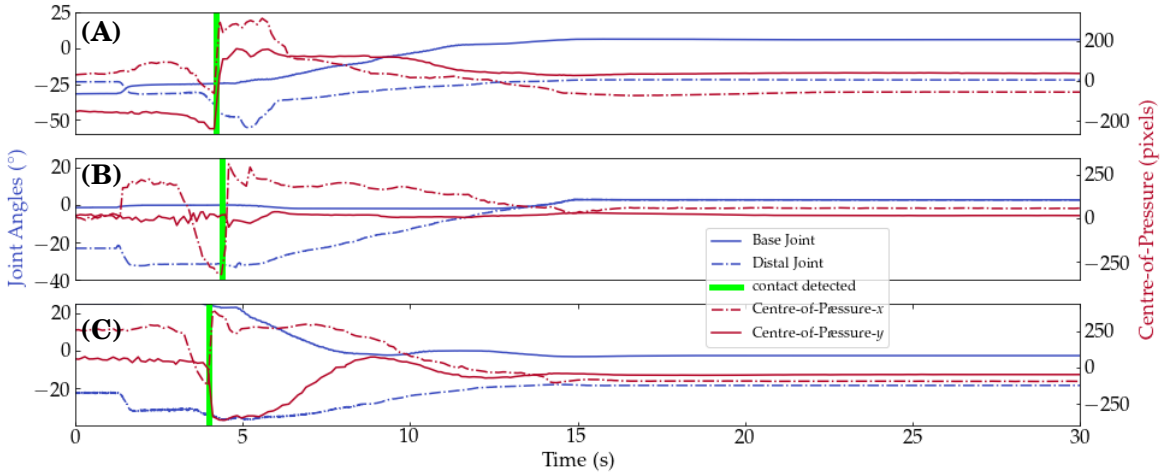

Figure 6.14: Plots of base and distal joint angles (blue) and *xy* centre-of-pressure (red) versus time, for fingers **A**, **B** and **C**, whilst grasping the Rubik's cube. Vertical green lines show when each finger detected contact.

respectively. This indicates a strong relationship between centre-of-pressure and joint angles.

In the case of unsuccessful grasps (please see supplementary video), the hand typically 'spills' the object. We attribute this to an 'over-adjustment'. The hand initially grasps the object with horizontal forces balanced across all fingertips. Joints are then adjusted. At some point, prior to time-out, adjustment moves the hand beyond a stable configuration; i.e., normal forces exerted by each fingertip become sufficiently un-balanced to push the object out of a grasp. We selected a time-out period of 10 s as it provided suitable time for the centre-of-pressure to approach zero in all cases. To avoid over-adjustment, the time-out could be lowered resulting in less adjustment but also less scope for the net force to push the object away from the grasp.

6.5 Concluding Remarks

In this chapter, we presented the integration of an established optical tactile sensing technology, the TacTip, with a three-fingered, commercial robot hand: the Shadow Modular Grasper. The sensors were tested for predicting roll and pitch relative to a flat surface and two generic shapes: a dome and an edge. Finally, we integrated tactile output with the hand control and demonstrated grasps on three real-world objects, using predicted roll and pitch angles to adjust the grasp for attaining greater contact surface areas at each finger-pad.

For predicting roll and pitch angles, three linear regression models were compared; simple, 2^{nd}- and 3^{rd}-order polynomials. The 2^{nd}-order model was the most suitable due to its balance of accuracy and generalisation. A clear limitation was identified when predicting orientations on an edge, where a singularity was observed for sensing perpendicular to the edge. This constraint is not limited to the presented hardware or method since it is produced by a physical singularity

observed in the raw tactile signal.

The Modular Grasper, equipped with the designed tactile fingertips, was able to successfully grasp and hold a Rubik's cube, Pringles can and mustard bottle. Additionally, the adjustment mechanism behaved as designed: to obtain a greater surface area of contact and a more centrally located pressure, both of which are likely to improve grasp quality.

The work presented in this chapter primarily contributes to the field of robot grasping and robot touch by demonstrating that the TacTip is easily modified for application with a fully actuated, three-fingered industrial robot hand, which adds to the other research-based robot hands which have previously seen modification with the addition of TacTip based devices [25, 178, 190]. Furthermore, we demonstrate how the high-resolution output of the TacTip enables extraction of high-level features that are easily interpreted and leveraged in a grasp-control framework without the need for large quantities of training data. This is contrasted with much of the literature on tactile hands and robot grasping, where data-driven approaches are employed with large quantities of training data. For example, Hyttinen et al. (2015) [294], Kwiatkowski et al. (2017) [206] and Cockburn et al. (2017) [207] demonstrate grasp stability prediction from low-resolution tactile data using data-driven approaches and extensive in-hand training with abstract hand-crafted tactile features in [294] and learned tactile features in [206] and [207]. And re-grasp policies are learned, as in Chebotar et al. (2016) [194], where abstract features were extracted from low resolution spatio-tempotal tactile data collected with three BioTac sensors attached to a three-fingered robot hand from 50 hours of training data and applied to a reinforcement learning algorithm.

Grasp success prediction [4] and a re-grasping procedure [192] was demonstrated by Calandra et al. (2017/2018) using a two fingered gripper equipped with two GelSight sensors. In the former study, a CNN was used to automatically learn tactile features from over 9000 manually labelled grasps and, in the latter, a CNN trained on GelSight images was used to predict grasp success of candidate adjustments from which the most promising actions were selected. The GelSight offers high-resolution sensing which enables easy extraction of meaningful tactile features, although the authors chose to automatically learn alternatives from extensive training data. Furthermore, the two-fingered gripper used in [4] and [192] is significantly less capable than the 9-DOF three-fingered hand used in this chapter, thus necessitating a distinct re-grasp rather than in-hand grasp adjustment as demonstrated in the present study.

In the present study, roughly half the trials were able to maintain a stable grasp throughout the entire adjustment phase. The primary cause of failure was 'over-adjustment': forces become un-balanced between the fingers resulting in the object being 'spilled'. We selected a time-out period of 10 s as it provided suitable time for the centre-of-pressure to approach zero in all cases. To avoid over-adjustment, the time-out could be lowered resulting in less adjustment but also less scope for the net force to push the object away from the grasp.

The results presented here are mainly descriptive on account of the preliminary nature of

this work. An important avenue for the progression of this work is to quantitatively asses the grasp quality before and after adjustment. Roa and Suárez (2015) [301] provide a comprehensive review of the measures proposed in the literature to evaluate grasp quality. The authors describe 24 different measures of grasp quality broadly categorised as considering contact locations on the object and/or hand configuration, both of which change throughout our grasp adjustment procedure. As pointed out in the Literature Review (Section 2.2.3), Chen et al. (2018) [239] also demonstrate that friction and slip monitoring are important for maintaining stable grasp. We suggest that for future work, as well as using metrics from [301] to asses grasp quality before and after adjustment, it may be beneficial to monitor some grasp quality metrics throughout the adjustment phase in order to modulate the degree of adjustment and thus balance aspects contributing to successful grasp; e.g., friction vs. balancing forces at contact locations. We propose that the grasp adjustment procedure could then be integrated into a high-level, modular controller, similar to the approach taken by Romano et al. [262].

Recently, OpenAI achieved in-hand manipulation with impressive levels of dexterity using deep reinforcement learning. To aid training, simulation was used, however, this is a challenge for touch, hence, this modality was omitted from their research. This highlights a need for learning-free approaches to using touch for application with any hand, without the need for re-training or simulation.

A key benefit of the presented approach for using touch is that, whilst it requires high-resolution tactile sensing, it can be easily applied to any robotic hand with the required degrees of freedom. This is owing to the intuitive output from the sensors (roll and pitch angle) which can be directly applied to joint angles of the gripper. As a consequence no prior training with the robot hand is required.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

This thesis sought to progress the field of tactile robotics by presenting a set of techniques based on theories of human touch. In particular, we proposed that multi-modality is a key aspect for robot touch based on an understanding of the range of schemes of sensory encoding in human tactile afferents [11, 23, 43]. To this end, we demonstrated two multi-modal tactile sensors, both based heavily on an established tactile sensor, the TacTip [13], and both of which are proposed to have sub-modalities which model human tactile channels. We also presented novel feature extraction techniques and suggest how these may be thought of as analogous to peripheral encoding schemes of tactile percepts in humans, to the benefit of artificial perception, in particular, texture discrimination. Finally, we demonstrated how the biomimetic TacTip sensor could be adapted for integration with an industrial robot hand and how tactile sensibility can aid robot grasping.

Here we summarise the findings of proceeding Chapters 3-6 (Section 7.1), we then discuss limitations of the research (Section 7.2) and finally we suggest potential future work towards developing human-like artificial robot touch (Section 7.2).

7.1 Conclusions

High-Resolution and High-Speed Dual Modal Optical Tactile Control (Chapter 3)
In this chapter, we presented a novel, dual-modal, optical tactile sensor, with a design based heavily on the TacTip [13]. The sensor has two modes of operation: high-speed mode (mode-HS) and high-resolution mode (mode-HR).

Mode-HS offered optical statistics at a significantly increased sample rate (2 kHz). We observed that in nature certain tactile modalities offer low-latency which is important for many

functions. Here, in particular, we relate an application of mode-HS to a possible function of FA or PC afferents which is contact detection [51]. We also suggest that mode-HS could offer function analogous to reflex action in animals, specifically for responding to over-exertion of normal force [31]. Here we proposed that this concept also has potential benefits for robots. We demonstrated how mode-HS could be used for contact depth modulation, the concept being that, in theory, a robot can quickly respond to excess pressure via the rapid sampling rate offered by mode-HS.

This chapter broadly demonstrated that cheap optical dual-modality is possible with a TacTip-like sensor. Importantly, the study also demonstrated that image processing was unnecessary for mode-HR which has implications for optical tactile sensing as it can potentially aid miniaturisation and reduce manufacturing cost through the use of smaller pixel arrays.

Artificial tactile afferents from markers and vibrations (Chapter 4)
Here we presented a number of hardware modifications to the TacTip, all of which were targeted towards developing methods of biologically-inspired artificial texture perception (Chapter 5). We augmented the TacTip with an artificial fingerprint and markers were arranged in a novel dense (~40 cm^{-2}) square (19x19) array. An additional hardware modality, the *vibrational channel*, was engineered via embedding a small electret microphone in the TacTip gel. The vibrational channel was proposed as an analogue of the natural PC channel.

From marker positions we presented two novel feature sets based on marker euclidean distance and marker speed which we suggest as artificial models of natural SA-I and FA afferent firing rates respectively. We were able to construct tactile images of these 'artificial afferents', which we suggest may provide spatial codes for tactile stimuli.

The feasibility of these artificial afferents as models of their natural counterparts was examined in a set of experiments. Notably, we compared 'spatial response profiles' (SRPs) of individual artificial SA-I and FA afferents, when systematically stimulated at positions along sets of aperiodic gratings, to similar stimulation in monkey SA-I and FA afferents [112]. Artificial afferents exhibited interesting similarities with their natural counterparts. In particular, enhanced edge sensitivity in SA-I afferents and modulation of sensitivity with reducing scales of stimulus dimensions in both afferent types. We use an interpretation of these SRPs presented by Phillips et al. [112] for estimating spatial resolution of afferents, which we observed as ~4 mm for both artificial SA-I and FA afferents. This is greater than natural SA-I afferents and roughly comparable with natural FA afferents.

We also aimed to understand the capacity for artificial SA-I, artificial FA and the vibrational channels to mediate information pertaining to texture via dynamic stimulation. We observed that the artificial fingerprint amplified vibrational amplitudes in all three channels. It was not obvious how SA-I and FA data might encode texture under dynamic touch. However, the frequency of peaks and their associated amplitudes observed in frequency spectra of the vibrational channel were closely correlated with bump spacing and bump diameter respectively, suggesting that 'harmonic structure' of induced vibrations may provide a viable cue for texture classification.

Artificial afferents applied to texture perception (Chapter 5)

Guided by our understanding of artificial tactile channels from the preceding chapter, here we presented a set of techniques for artificial texture discrimination using artificial encoding schemes and vibrational cues inspired by those believed to be employed in human texture perception. As well as developing practical solutions to an ongoing area of research in tactile robotics, we also wished to use the results of this chapter to further understand how well our artificial tactile channels model their natural analogues.

Our analysis of the spatial resolution of artificial SA-I and FA afferents from the preceding chapter suggested that spatial resolution when statically stimulated (pressing) with raised bump stimuli would be limited to bump spacings <~4 mm. In fact, artificial SA-I tactile images were good predictors for textures over the entire range, as demonstrated by the capacity of CNNs to decode texture class over the entire range. This suggested that *(i)* tactile images of SA-I can spatially encode texture and *(ii)* some non-spatial cue was mediated through SA-I tactile images. We did however observe that accuracy was improved for textures with bump spacing >~4 mm, suggesting that above this threshold additional spatial cues were available. This is somewhat indicative of static texture perception in humans, the performance of which has been shown to degrade for increasingly fine textures [57].

Data collected via sliding stimuli relative to the sensor were dominated by shear, suggesting that frictional rather than spatial cues would be viable predictors of texture. Under sliding stimulation, the addition of a temporal dimension and use of convLSTM to extract spatio-temporal features improved accuracy for all models trained on artificial FA tactile images but did not affect performance when using SA-I data. This result may be indicative of properties of natural SA-I and FA afferents since neurophysiological readings have shown that FA fibres respond more robustly than SA-I to dynamic stimulation with a range of natural textures [129].

From the previous chapter we identified that harmonic structure of induced vibrations collected with the vibrational channel could provide a viable cue for texture discrimination. Furthermore, in this chapter, we identified that harmonic structure was speed invariant. This compelled us to use FFT samples for texture prediction in order to mimic the observed speed invariance of human texture perception [131]. We used 1D-CNNs to learn relevant structural features from the FFT samples. However, despite the availability of speed-invariant harmonic structure, this method was unable to learn features that were robust to speed. To counter this, we provided an augmented training set, where additional speeds were simulated, which greatly improved performance up to 100 % accuracy on some test sets.

Application to robot hands (Chapter 6)

Here we demonstrated integration of the biomimetic TacTip sensor with an industrial, 3-fingered robot hand, the Shadow Modular Grasper [27]. We successfully miniaturised the TacTip to create a set of three tactile fingertips without the need for any additional expensive equipment; e.g., miniature cameras.

We demonstrated how a high-level tactile feature (centre-of-pressure) can be used with simple regression techniques to predict roll and pitch of the tactile fingertip relative to a stimulus surface. We integrated predictions of roll and pitch with an existing grasp controller, by mapping roll and pitch directly to relevant finger joint angles, and demonstrated how these predictions could be used for adjusting and simultaneously maintaining the grasp in order to increase surface area of contact on three real-world objects: a Rubik's cube, a Pringles can and a mustard bottle.

It was suggested that increasing surface area of contact improves frictional forces and therefore grasp quality. In practice, half the trails were able to maintain a stable grasp throughout the entire adjustment phase. The primary cause of failure was 'over-adjustment' as forces become un-balanced between the fingers resulting in the object being 'spilled'. We propose that an effective amount of adjustment may depend on the object being grasped and by adapting the adjustment period or joint limits the success rate may be improved.

To summarise, we have demonstrated that the biomimetic TacTip sensor can be endowed with additional and practical sub-modalities. We have shown that derivations of TacTip marker positions provide viable models for natural SA-I and FA afferents through comparison with neurophysiological readings and that these artificial afferents can be used with conventional machine learning techniques to discriminate texture. The performance of these artificial models was also indicative of human texture perception. We also demonstrated that high-resolution biomimetic optical tactile sensing provided by the TacTip is a viable solution for integration with robot hands. A significant contribution of this thesis is the concept of considering transduction and feature engineering as analogous to neural codes in humans, which we presented in the literature review. We feel that this guided our solution for robot texture perception and can have significant impact in other areas of robot touch.

7.2 Limitations

A range limitations were highlighted throughout the thesis. For the purpose of consolidation, however, we summarise only the limitations with broader applicability here.

Manufacturing of most research materials was restricted to 3D-printing. Whilst this is a fantastic tool for rapid prototyping, there is an associated cost in terms of print resolution (depending on the model of printer). This presented two main limitations. Firstly, the maximum density of markers in the 3D-printed tip (Chapter 4). This was roughly half that of SA-I afferents in the human fingertip. By increasing marker density we presumably would have a better model for type-I afferents. Secondly, we were limited on the 'fineness' of textured surface stimuli used in the texture discrimination task (Chapter 5), where the closest achievable bump spacing was ~0.4 mm. We hypothesised that, in accordance with theories of human texture perception, SA-I and FA afferents may be ineffective for discrimination of fine textures, thus demonstrating the need for

the PC channel. In practice, we observed that the drop in performance for fine textures was not appreciable, particularly when using artificial FA afferents stimulated with sliding. We believe for finer textures, the ability for the TacTip to transduce viable spatio-temporal codes may be limited both due to spatial resolution of the TacTip skin and sample rate of the camera. In this case we may see greater disparity between the performance of vibrational channel and that of artificial SA-I and FA afferents.

Also a consequence of 3D-printing, the detail exhibited in spatial response profiles (SRPs) of the central artificial afferent (Chapter 4) was considerably less than that of their natural counterpart. As outlined in Chapter 4, we believe this was a due to skin dynamics, where the 3D-printed skin was unable to conform to small stimulus gaps. We suggest that with a more conformable skin the SRPs may show more detail, which is indicative of better spatial resolution.

The approach taken to modelling natural SA-I and FA afferents was to assume a rate code (Chapter 4): essentially, output at each frame was related to spike count within an encoding window of a neuron. The assumption of a fundamental rate code is essential for the TacTip because it does not produce spikes, so in a sense this is a limitation. An alternative approach would be to use a neuromorphic sensor capable of producing spike based output; i.e., stimuli may be encoded in precise spike timing. Whilst this is appealing from an academic standpoint, neuromorphic sensors are considerably more expensive than the TacTip. Furthermore, to leverage the spike timing of neuromorphic signals requires complex decoders such as spiking neural networks (SNNs) [302], where there are many open questions in their application, instead we were able to employ standard machine learning techniques.

7.3 Future Work

The work presented within this thesis, whilst drawn to a logical conclusion, has natural extensions for which we believe can further progress the field.

Whilst we attempted to model a neural code of natural SA-I and FA afferents using markers (Chapter 4), we did not do the same for natural PC afferents and the novel vibrational channel, which was proposed to be analogous to the PC system. The data collected with this channel is a pure representation of the stimulus vibrations, so to propose a model of afferent firing would require extracting some cue (e.g., amplitude, frequency or some more complex combination) from the data, choosing an encoding scheme (e.g., rate or temporal) and generating a model to map the cue to the chosen artificial code. We stated that, as a novel tactile channel, this was beyond the scope of the work and chose instead to focus only on looking at viable tactile cues for speed invariant texture perception. We find that generating an artificial model for PC afferents is an interesting and logical next step towards creating a full model of human peripheral somatosensory system.

We identified that the harmonic structure of induced vibrations was a viable cue for speed invariant texture perception (Chapter 5) and studies have shown that the frequency response of skin vibrations was reflected in the frequency of PC afferent response when fingers were scanned with natural textures. This seems to confirm the plausibility of harmonic structure as a viable cue, if stimuli are encoded temporally; i.e., within precise timing of individual spikes. To leverage a pure temporal code would require a spiking output; i.e., some sort of neuron model which is a potential future direction of the work.

In this thesis, we chose to use controlled artificial textures (raised bumps; Chapters 4 and 5) because we wished to gain an understanding of the fundamentals of texture perception, e.g skin dynamics and limits on resolution, for the purpose of identifying viable cues and developing artificial encoding schemes. However, it is common within the robotics community, and indeed more applicable to the real-world, to use natural textures. Having gained an understanding of the underlying mechanisms for artificial texture perception, a logical progression is to test our techniques on natural textures and evaluate how well the methods transfer.

We sought to gain an understanding of how well our artificial perceptual system reflected tactile perception in humans (Chapter 5). There were limits on the extent to which we were able to draw comparison because of a fundamental difference in how the tasks were performed. We chose to use a classical robotics test, a classification task, whereas, studies into human perception use psychophysical tasks. For example, two-alternative forced choice (2AFC), where the participant undergoes a number of trials and on each trial is required to identify a target stimulus out of two presentations. Between each trial the intensity difference between the stimuli in each presentation is adjusted [85]. The proportion of correct responses as a function of intensity difference, known as the psychometric function, is a key tool in assessing perceptual performance. We wish to generate a framework for carrying out such tasks on a robot, so that robot-psychometric functions can be produced and directly compared to that of humans.

In this thesis, we have performed studies on three relatively distinct pieces of hardware: a dual-model tactile sensor for pain and reflex sensing (Chapter 3), a modified TacTip for modelling discriminative tactile afferents (Chapters 4 and 5) and an industrial hand with biomimetic tactile fingertips (Chapter 6). We feel, therefore, an exciting and logical progression of the work presented in this thesis would be the integration of these systems into a human inspired, multi-modal tactile robot hand. We stated in the introduction of the thesis that we believe “a holistic approach to artificial touch is required if robots are to achieve human levels of dexterity and control”. Certainly, this thesis has made progress towards this end, and a single system which brings its concepts together would be a valuable extension of the work and a significant contribution to the field.

APPENDIX A

A.1 Robot Nociception (Chapter 3)

A.1.1 Interfacing

A.1.1.1 Arduino Nano - ADNS-3080

Fig. A.1 shows a diagram of the hardware interfacing between Arduino and ADNS-3080.

The four wires are: clock (SCLK), chip-select (NCS), master-out-slave-in (MOSI) and master-in-slave-out (MISO). The Arduino powers the ADNS-3080 via 5V and ground lines. Reseting the ADNS-3080 is available via a reset line. SCLK is a a synchronisation reference driven by the master device (Arduino) it is set to a frequency of 500 kHz. In general, NCS is used for selecting a slave device. Here we have only one slave device, the ADNS-3080. In this case, NCS is used to enable the serial port by setting the line LOW. All transactions are initiate by a change in NCS from HIGH to LOW and are terminated by a change from LOW to HIGH.

For write operations, 2 bytes are driven by the Arduino on the MOSI line. The first byte consists of a '1' as the MSB, which signifies a write operation, and a 7-bit address for the ADNS-3080 register. The second byte contains the data.

For read operations, e.g. recording sensor data, an initial byte is driven by the Arduino to the ADNS-3080 on the MOSI line containing a '0' as the MSB, signifying a read operation, and a 7-bit address for the register which is to be read. After a short delay, $t_{SRAD} = 50\mu s$, the ADNS-3080 responds by driving the relevant data on the MISO line [246].

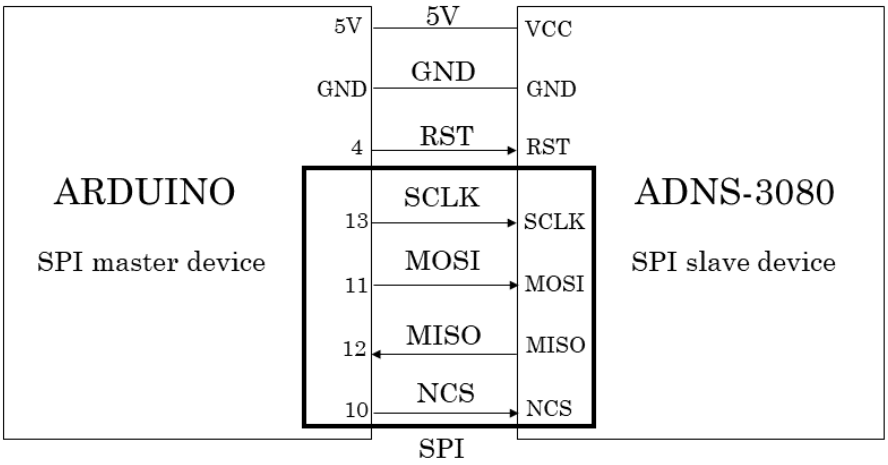

Figure A.1: Diagram of data (SPI, enclosed in bold box) and power connections between the Arduino Nano and ADNS-3080.

Diagrams of write and read operations are shown in Figs. A.2 and A.3 respectively.

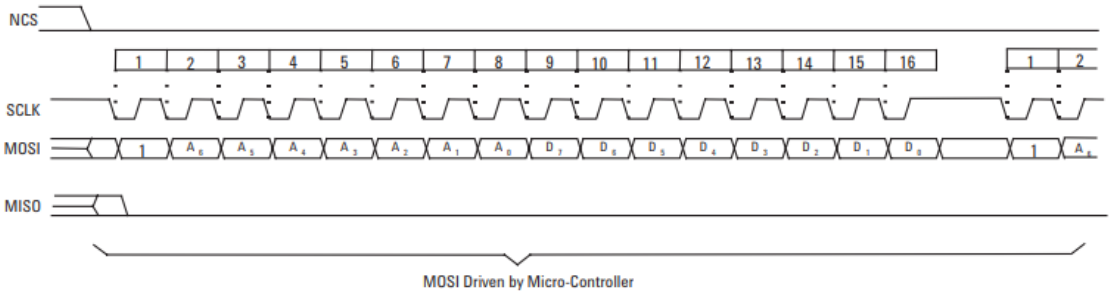

Figure A.2: Diagram of write operation with synchronous serial port on ADNS-3080 [246].

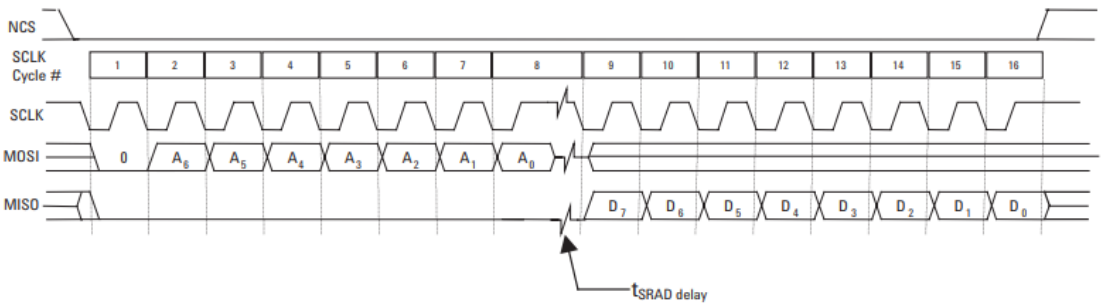

Figure A.3: Diagram of write operation with synchronous serial port on ADNS-3080 [246].

A.1.1.2 Sensor - PC

For the purpose of conceptualising hardware interfaces, the tactile sensor is considered to be simply comprised of the ADNS-3080 and the Arduino Nano (see Fig. A.4).

The tactile sensor is interfaced with an external PC through an asynchronous serial transmission via the Arduino's serial port and a virtual COM port on the PC. In practice, an FTDI chip [303] on the Arduino converts data that comes from its main processor from serial universal-asynchronous-receiver-transmiter (UART) to USB.

The USB protocol is used to transmit data to a virtual COM port on the PC. This virtual COM port allows applications to interface with the Arduino as a standard serial device so applications written for either the PC or the Arduino may treat communication as standard UART. Buffers exist on both the PC and the Arduino. When applications read or write serial data they are put on or taken off these buffers respectively, in a first in-first out (FIFO) manner. Data is asynchronously piped from one buffer to the other where sending of each bit is triggered via an interrupt handler. The rate at which this interrupt handler is triggered can be adjusted via the baud-rate.

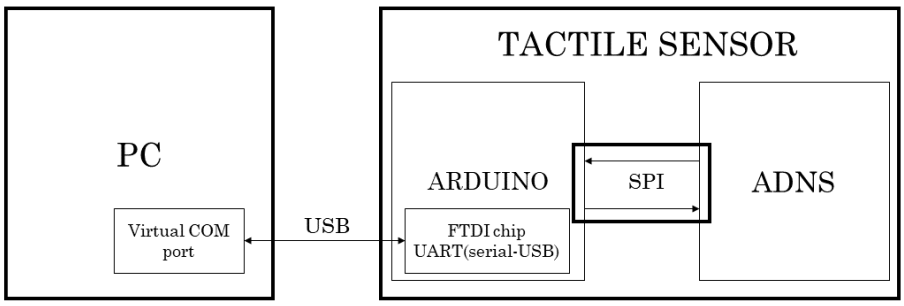

Figure A.4: Diagram of overall hardware interface. The tactile sensor (left), comprised of the Arduino nano and the ADNS-3080 is interfaced with the PC (right) via USB.

A.1.2 ADNS-3080 Burst Mode

The Motion_Burst_Mode register (0x50) is addressed and the ADNS-3080 responds after a short time ($t_{SRAD\text{-}MOT}$ = 75 μs) with the contents of the Motion, Delta_X, Delta_Y, SQUAL, Shutter_Upper, Shutter_Lower and Maximum_Pixel registers in that order [246]. A diagram of this approach is provided in Fig. A.5.

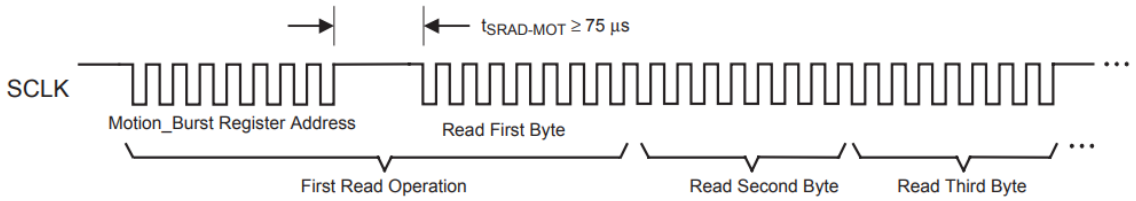

Figure A.5: Diagram of read-write method for using the ADNS-3080 motion burst mode [246].

A.2 A Novel, Multi-Modal Tactile Sensor (Chapter 4)

A.2.1 Comparison with Physiological Data: Supplementary Figures

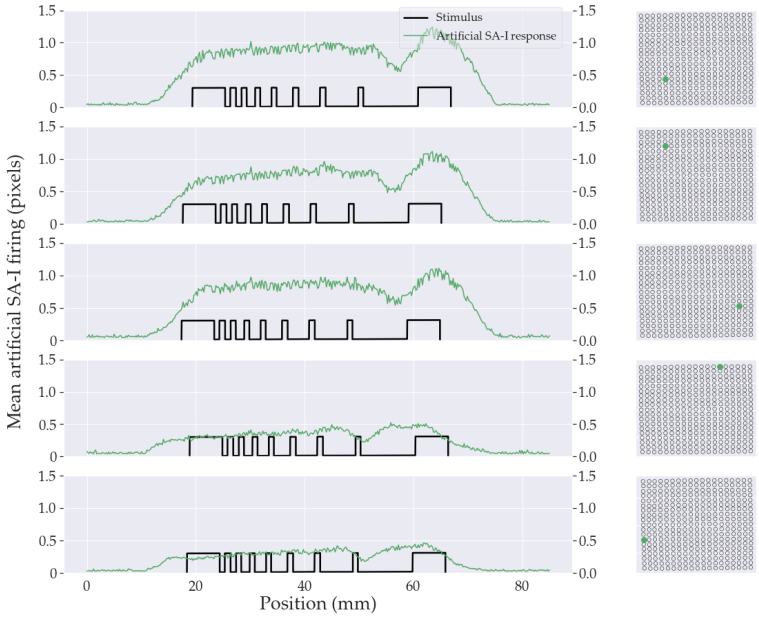

Figure A.6: *Left*: Spatial response profiles (SRPs) for a non-centrally located artificial SA-I afferents of the smooth tip (green) collected on grating **A** (black). *Right*: position of selected artificial SA-I afferent (green).

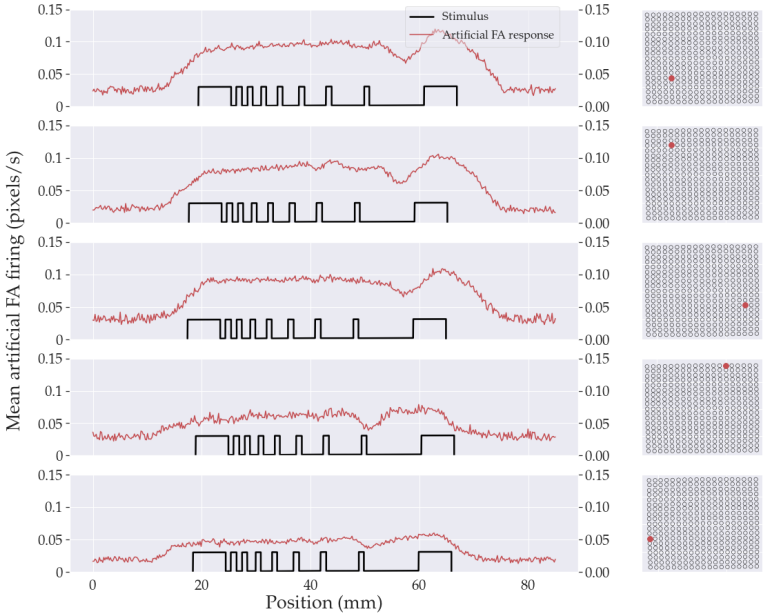

Figure A.7: *Left*: SRPs for a non-centrally located artificial FA afferents of the smooth tip (red) collected on grating **A** (black). *Right*: position of selected artificial FA afferent (red).

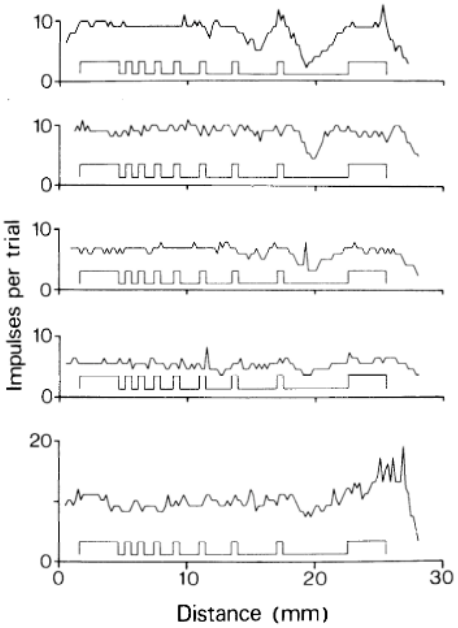

Figure A.8: SRPs of discharge from five seperate FA afferents innervating a monkey's fingerpad when indented at distinct locations across the displayed grating (below). From Phillips and Johnson (1981) [112].

BIBLIOGRAPHY

[1] Z. Kappassov, J.-A. Corrales, and V. Perdereau.
Tactile sensing in dexterous robot hands: a review.
Robotics and Autonomous Systems, 74:195 – 220, 2015.

[2] D. W. Hillis.
A high-resolution imaging touch sensor.
International Journal of Robotics Research, pages 1–2, 1982.

[3] K. Goldberg and R. Bajcsy.
Active touch and robot perception.
01 1984.

[4] R. Calandra, A. Owens, M. Upadhyaya, W. Yuan, J. Lin, E. H. Adelson, and S. Levine.
The feeling of success : Does touch sensing help predict grasp outcomes ?
In *1st Conference on Robot Learning (CoRL 2017), Mountain View, United States*, 2017.

[5] Q. li, C. Schuermann, R. Haschke, and H. Ritter.
A control framework for tactile servoing.
06 2013.

[6] N. F. Lepora, A. Church, C. De Kerckhove, R. Hadsell, and J. Lloyd.
From pixels to percepts: Highly robust edge perception and contour following using deep learning and an optical biomimetic tactile sensor.
CoRR, abs/1812.02941, 2018.

[7] D. Xu, G. E. Loeb, and J. A. Fishel.
Tactile identification of objects using bayesian exploration.
In *2013 IEEE International Conference on Robotics and Automation*, pages 3056–3061, May 2013.

[8] W. Yuan, C. Zhu, A. Owens, M. A. Srinivasan, and E. H. Adelson.
Shape-independent hardness estimation using deep learning and a gelsight tactile sensor.
In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 951–958, May 2017.

[9] V. E. Abraira and D. D. Ginty.
The sensory neurons of touch.
Neuron, 79(4):618 – 639, 2013.

[10] R. L Klatzky and S. J. Lederman.
Toward a computational model of constraint-driven exploration and haptic object identification.
Perception, 22(5):597–621, 1993.

[11] H. Saal and S. Bensmaïa.
Touch is a team effort: interplay of submodalities in cutaneous sensibility.
Trends in Neurosciences, 37:689–697, 2014.

[12] C. Chorley, C. Melhuish, T. Pipe, and J. Rossiter.
Development of a tactile sensor based on biologically inspired edge encoding.
In *2009 International Conference on Advanced Robotics*, pages 1–6, 2009.

[13] B. Ward-Cherrier, N. Pestell, L. Cramphorn, B. Winstone, M. E. Giannaccini, J. Rossiter, and N. F. Lepora.
The tactip family: Soft optical tactile sensors with 3d-printed biomimetic morphologies.
Soft Robotics, 2018.

[14] C. Allègre, G. Manhès, and C. Gopel.
The age of the earth.
Geochimica et Cosmochimica Acta, 59(8):1445 – 1456, 1995.

[15] C. Darwin.
On the Origin of Species by Means of Natural Selection.
Murray, London, 1859.
or the Preservation of Favored Races in the Struggle for Life.

[16] J. M. Harkness.
In appreciation6a lifetime of connections: Otto herbert schmitt, 1913 - 1998.
Physics in Perspective, 4:456–490, 2002.

[17] J. Vincent, O. Bogatyreva, N. Bogatyrev, A. Bowyer, and A.-K. Pahl.
Biomimetics: Its practice and theory.
Journal of the Royal Society, Interface / the Royal Society, 3:471–82, 09 2006.

[18] L. D. Paulson.
Biomimetic robots.
Computer, 37(9):48–53, 2004.

[19] D. Acemoglu and P. Restrepo.
Robots and jobs: Evidence from us labor markets.
Journal of Political Economy, 0, 0.

[20] T. Mitzner, T. Chen, C. Kemp, and W. Rogers.
Identifying the potential for robotics to assist older adults in different living environments.
International journal of social robotics, 6:213–227, 04 2014.

[21] K. Elizabeth E. Broadbent, H. I. Kuo, Y. Lee, J. Rabindran, N. Kerse, R. Stafford, and B. A. MacDonald.
Attitudes and reactions to a healthcare robot.
Telemedicine and e-Health, 16(5):608–613, 2010.

[22] G. A. Bekey.
On autonomous robots.
The Knowledge Engineering Review, 13(2):143–146, 1998.

[23] S. J. Lederman and R. L. Klatzky.
Haptic perception: A tutorial.
Attention, Perception, & Psychophysics, 71(7):1439–1459, Oct 2009.

[24] U. Martinez-Hernandez, G. Metta, T. J. Dodd, T. J. Prescott, L. Natale, and N. F. Lepora.
Active contour following to explore object shape with robot touch.
In *2013 World Haptics Conference (WHC)*, pages 341–346, April 2013.

[25] B. Ward-Cherrier, L. Cramphorn, and N. F. Lepora.
Tactile manipulation with a tacthumb integrated on the open-hand m2 gripper.
IEEE Robotics and Automation Letters, 1(1):169–175, Jan 2016.

[26] N. F. Lepora and B. Ward-Cherrier.
Superresolution with an optical tactile sensor.
In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2686–2691, 2015.

[27] Shadow Robot Company agile grasper documentation.
`https://agile-grasper.readthedocs.io/en/latest/`.
Accessed: 2018-09-07.

[28] E.R. Kandel, T.M. Jessell, J.H. Schwartz, S.A. Siegelbaum, and A.J. Hudspeth.
Principles of Neural Science, Fifth Edition.
Principles of Neural Science. McGraw-Hill Education, 2013.

[29] B. Hile.

Ionic Channels of Excitable Membranes.
Ionic Channels of Excitable Membranes. Cell Press, 1992.

[30] R. S. Johansson and A. B. Vallbo.
Tactile sensory coding in the glabrous skin of the human hand.
Trends in Neurosciences, 6:27 – 32, 1983.

[31] A. I. Basbaum, D. M. Bautista, G. Scherrer, and D. Julius.
Cellular and molecular mechanisms of pain.
Cell, 139:267 – 84, 2009.

[32] F. McGlone, J. Wessberg, and H. Olausson.
Discriminative and affective touch: Sensing and feeling.
Neuron, 82(4):737 – 755, 2014.

[33] P. Grigg.
Peripheral neural mechanisms in proprioception.
Journal of Sport Rehabilitation, 3(1):2–17, 1994.

[34] I. A. Boyd.
The isolated mammalian muscle spindle.
Trends in Neurosciences, 3(11):258 – 265, 1980.

[35] R. S. Dahiya, G. Metta, M. Valle, and G. Sandini.
Tactile sensing; from humans to humanoids.
IEEE Transactions on Robotics, 26(1):1–20, 2010.

[36] J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson.
Handbook of psychophysiology.
2007.

[37] A. Iggo and K. H. Andres, editors.
Morphology of Cutaneous Receptors.
Springer Berlin Heidelberg, Berlin, Heidelberg, 1973.

[38] M. Ksahara M. R. Miller, H. J. Ralston.
Chapter 1 - the pattern of cutaneous innervation of the huan hand, foot and breast.
In W, editor, *Cutaneous Innervation*, pages 1 – 47. Pergamon, 1960.

[39] G. J. Gerling and G. W. Thomas.
The effect of fingertip microstructures on tactile edge perception.
In *First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics Conference*, pages 63–72, 2005.

[40] G. J. Gerling.
Sa-i mechanoreceptor position in fingertip skin may impact sensitivity to edge stimuli.
Applied Bionics and Biomechanics, 7:19–29, 2010.

[41] N. Cauna.
Nature and functions of the papillary ridges of the digital skin.
The Anatomical record, 119 4:449–68, 1954.

[42] R. Johansson and AB Vallbo.
Tactile sensibility in the human hand: Relative and absolute density of four types of mechanoreceptive units in glabrous skin.
The Journal of physiology, 286:283–300, 02 1979.

[43] K. O. Johnson and J. R. Phillips.
Tactile spatial resolution. I. two-point discrimination, gap detection, grating resolution, and letter recognition.
Journal of neurophysiology, 46(6):1177–1192, 1981.

[44] G. Westling and R. S. Johansson.
Responses in glabrous skin mechanoreceptors during precision grip in humans.
Experimental Brain Research, 66(1):128–140, 1987.

[45] R. Johansson and A.B. Vallbo.
Detection of tactile stimuli. thresholds of afferent units related to psychophysical thresholds in the human hand.
The Journal of physiology, 297:405–22, 01 1980.

[46] E. Ergen and B. Ulkar.
Chapter 18 - proprioception and coordination.
In Walter R. Frontera, Stanley A. Herring, Lyle J. Micheli, Julie K. Silver, and Timothy P. Young, editors, *Clinical Sports Medicine*, pages 237 – 255. W.B. Saunders, Edinburgh, 2007.

[47] C. F. Bolton, R. K. Winkelmann, and Peter J Dyck.
A quantitative study of meissner's corpuscles in man.
Transactions of the American Neurological Association, 89:190–192, 1964.

[48] M. Knibestöl.
Stimulus-response functions of rapidly adapting mechanoreceptors in human glabrous skin area.
The Journal of physiology, 232:427–52, 08 1973.

[49] M. Muniak, S. Ray, S. Hsiao, J. Frank J. Dammann, and S. Bensmaïa.
The neural coding of stimulus intensity: Linking the population response of mechanoreceptive afferents with psychophysical behavior.
Journal of Neuroscience, 27(43):11687–11699, 2007.

[50] Y. Roudaut, A. Lonigro, B. Coste, J. Hao, P. Delmas, and M. Crest.
Touch sense.
Channels, 6(4):234–245, 2012.

[51] R. S. Johansson and A. B. Vallbo.
Detection of tactile stimuli. thresholds of afferent units related to psychophysical thresholds in the human hand.
The Journal of Physiology, 297(1):405–422, 1979.

[52] S. Bolanowski and L. Pawson.
Organization of meissner corpuscles in the glabrous skin of monkey and cat.
Somatosensory & motor research, 20:223–31, 09 2003.

[53] J. Hoffmann, A. Montag, and N. Dominy.
Meissner corpuscles and somatosensory acuity: The prehensile appendages of primates and elephants.
The Anatomical Record Part A Discoveries in Molecular Cellular and Evolutionary Biology, 281:1138–47, 11 2004.

[54] R. Johansson.
Tactile sensibility in the human hand: Receptive field characteristics of mechanoreceptive units in the glabrous skin area.
The Journal of physiology, 281:101–25, 09 1978.

[55] R. H. LaMotte and V. B. Mountcastle.
Capacities of humans and monkeys to discriminate vibratory stimuli of different frequency and amplitude: a correlation between neural events and psychological measurements.
Journal of neurophysiology, 38(3):539–559, 1975.

[56] R.S. Johansson, U. Landström, and R. Lundström.
Responses of mechanoreceptive afferent units in the glabrous skin of the human hand to sinusoidal skin displacements.
Brain Research, 244(1):17 – 25, 1982.

[57] M. Hollins and S. R. Risner.
Evidence for the duplex theory of tactile texture perception.
Perception and Psychophysics, 62(4):695–705, 2000.

[58] S. Bensmaïa and M. Hollins.
Pacinian representations of fine surface texture.
Perception and Psychophysics, 67(5):842–854, 2005.

[59] J. Scheibert, S. Leurent, A. Prevost, and G. Debrégeas.
The role of fingerprints in the coding of tactile information probed with a biomimetic sensor.
Science, 323(5920):1503–1506, 2009.

[60] S. Gandevia.
Kinesthesia: Roles for Afferent Signals and Motor Commands.
01 2011.

[61] T. C. Tuthill and E. Azim.
Proprioception.
Current Biology, 28(5):R194–R203, 2018.

[62] E. Marani and E. A.J.F. Lakke.
Chapter 4 - peripheral nervous system topics.
In J K. Mai and G. Paxinos, editors, *The Human Nervous System (Third Edition)*, pages 82 – 140. Academic Press, San Diego, third edition edition, 2012.

[63] J. C. Moore.
The golgi tendon organ: a review and update.
The American journal of occupational therapy : official publication of the American Occupational Therapy Association, 38(4):227âĂŤ236, April 1984.

[64] D. Fitzpatrick L. C. Katz A.-S. LaMantia J. O. McNamara D. Purves, G. J. Augustine, editor.
Neuroscience, 2nd edition.
Sunderland (MA): Sinauer Associates, 2001.

[65] S. Garrison, A. Dietrich, and C. Stucky.
Trpc1 contributes to light-touch sensation and mechanical responses in low-threshold cutaneous sensory neurons. j neurophysiol 107:913-922.
Journal of neurophysiology, 107:913–22, 11 2011.

[66] I. Croy, I. Sehlstedt, H. B. Wasling, R. Ackerley, and H. Olausson.
Gentle touch perception: From early childhood to adolescence.
Developmental Cognitive Neuroscience, 35:81 – 86, 2019.
Social Touch: A new vista for developmental cognitive neuroscience?

[67] J. Liljencrantz and H. Olausson.
Tactile c fibers and their contributions to pleasant sensations and to tactile allodynia.
Frontiers in Behavioral Neuroscience, 8:37, 2014.

[68] J. R. Augustine.
Circuitry and functional aspects of the insular lobe in primates including humans.
Brain Research Reviews, 22(3):229 – 244, 1996.

[69] L. Paninski, J. Pillow, and J. Lewi.
Statistical models for neural encoding, decoding, and optimal stimulus design.
In Paul Cisek, Trevor Drew, and John F. Kalaska, editors, *Computational Neuroscience: Theoretical Insights into Brain Function*, volume 165 of *Progress in Brain Research*, pages 493 – 507. Elsevier, 2007.

[70] M. Szwed, K. Bagdasarian, and E. Ahissar.
Encoding of vibrissal active touch.
Neuron, 40(3):621 – 630, 2003.

[71] J. W. Pillow, L. Paninski, V. J. Uzzell, E. P. Simoncelli, and E. J. Chichilnisky.
Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model.
Journal of Neuroscience, 25(47):11003–11013, 2005.

[72] C. R. Holdgraf, J. W. Rieger, and C. Micheli.
Encoding and decoding models in cognitive electrophysiology.
Frontiers in Systems Neuroscience, 11:1662–5137, 2017.

[73] F. Theunissen and J. P. Miller.
Temporal encoding in nervous systems: a rigorous definition.
Journal of computational neuroscience, 2:149–62, 07 1995.

[74] P. Dayan and L. F. Abbott.
Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems.
The MIT Press, 2005.

[75] M. A. Srinivasan, J. M. Whitehouse, and R. H. LaMotte.
Tactile detection of slip: surface microgeometry and peripheral neural codes.
Journal of Neurophysiology, 63(6):1323–1332, 1990.

[76] E. Brodie and H. E. Ross.
Sensorimotor mechanisms in weight discrimination.
Perception & Psychophysics, 36(5):477–481, Sep 1984.

[77] Y. Iwamura and M. Tanaka.
Postcentral neurons in hand region of area 2: their possible role in the form discrimination of tactile objects.
Brain Research, 150(3):662 – 666, 1978.

[78] L. G. Ungerleider and J. V. Haxby.
'what' and 'where' in the human brain.
Current Opinion in Neurobiology, 4(2):157 – 165, 1994.

[79] C. Reed, R. Klatzky, and E. Halgren.
What vs. where in touch: An fmri study.
NeuroImage, 25:718–26, 05 2005.

[80] G.A. Gescheider.
Psychophysics: The Fundamentals.
Taylor & Francis, 2013.

[81] V. B. Mountcastle, R. Lamotte, and G. Carli.
Detection thresholds for stimuli in humans and monkeys: comparison with threshold events in mechanoreceptive afferent nerve fibers innervating the monkey hand.
Journal of neurophysiology, 35:122–36, 02 1972.

[82] K. O. Johnson, T. Yoshioka, and F. Vega-Bermudez.
Tactile functions of mechanoreceptive afferents innervating the hand.
Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society, 17:539–58, 12 2000.

[83] H. Z. Tan, X. Dong P., and N. I. Durlach.
Manual resolution of length, force, and compliance.
Advances in robotics, 42:13–18, 1992.

[84] H. Tan, H. Kazerooni, J. E. Colgate, B. D. Adelstein, Book No. H, Hong Z. Tan, Nathaniel I. Durlach, Yun Shao, and Min Wei.
Manual resolution of compliance when work and force cues are minimized.
Advances in robotics, 49:99 – 104, 1993.

[85] D. M. Green and J. A. Swets.
Signal detection theory and psychophysics.
Wiley, New York, 1966.

[86] M. A. Srinivasan and R. H. LaMotte.
Tactual discrimination of softness.
Journal of Neurophysiology, 73(1):88–101, 1995.

[87] R. H. Lamotte.
Softness discrimination with a tool.
Journal of neurophysiology, 83 4:1777–86, 2000.

[88] E.H. Weber.
E. H. Weber: the sense of touch.
Experimental Psychology Series. Academic Press for Experimental Psychology Society, 1978.

[89] W. D. Augustus.
The sense of effort: an objective study.
Brain, 14(2-3):179–249, 04 1891.

[90] A. W. Goodwin, V. G. Macefield, and J. W. Bisley.
Encoding of object curvature by tactile afferents from human fingers.
Journal of neurophysiology, 78 6:2881–8, 1997.

[91] J. R. Phillips, K. O. Johnson, and S. Hsiao.
Spatial pattern representation and transformation in monkey somatosensory cortex.
Proceedings of the National Academy of Sciences of the United States of America, 85:1317–21, 03 1988.

[92] M. A. Srinivasan and R. H. LaMotte.
Tactile discrimination of shape: responses of slowly and rapidly adapting mechanoreceptive afferents to a step indented into the monkey fingerpad.
Journal of Neuroscience, 7(6):1682–1697, 1987.

[93] R. H. LaMotte and M. A. Srinivasan.
Responses of cutaneous mechanoreceptors to the shape of objects applied to the primate fingerpad.
Acta Psychologica, 84(1):41 – 51, 1993.
Tactile Pattern Recognition.

[94] A. L. Skinner, C. Kent, J. M. Rossiter, C. P. Benton, M G. M. Groen, and J. Noyes.
On the edge: Haptic discrimination of edge sharpness.
PLoS ONE, 2013.

[95] J. Park, A. J. Doxon, W. R. Provancher, D. E. Johnson, and H. Z. Tan.
Haptic edge sharpness perception with a contact location display.
IEEE Transactions on Haptics, 5(4):323–331, 2012.

[96] Chris Kent, Andrew L Skinner, Carys Weeds, and Christopher P Benton.
Proximal-distal, not medial-lateral, movement across an edge increases discrimination of edge sharpness.
PERCEPTION, 2014.

[97] R. J. Seitz, P. E. Roland, C. Bohm, T. Greitz, and S. Stone-Elander.
Somatosensory discrimination of shape: Tactile exploration and cerebral activation.
European Journal of Neuroscience, 3(6):481–492, 1991.

[98] B. T. O'Sullivan, P. E. Roland, and R. Kawashima.
A pet study of somatosensory discrimination in man. microgeometry versus macrogeometry.
European Journal of Neuroscience, 6(1):137–148, 1994.

[99] M. R. Stoesz, M. Zhang, V. D. Weisser, S. C. Prather, H. Mao, and K. Sathian.
Neural networks active during tactile form perception: common and differential activity during macrospatial and microspatial tasks.
International Journal of Psychophysiology, 50(1):41 – 49, 2003.
Current findings in multisensory research.

[100] A. Bodegård, S. Geyer, C. Grefkes, K. Zilles, and P. E. Roland.
Hierarchical processing of tactile shape in the human brain.
Neuron, 31(2):317 – 328, 2001.

[101] D. Filingeri.
Neurophysiology of skin thermal sensations.
Comprehensive Physiology, 6, 07 2016.

[102] L. A. Jones and S. J. Lederman.
Human hand function.
Oxford University Press, New York, NY, US, 2006.

[103] H. Hensel and A. Iggo.
Analysis of cutaneous warm and cold fibres in primates.
Pflügers Archiv, 329(1):1–8, Mar 1971.

[104] I. Darian-Smith, K. O. Johnson, C. LaMotte, Y. Shigenaga, P. Kenins, and P. Champness.
Warm fibers innervating palmar and digital skin of the monkey: Responses to thermal stimuli.
Journal of neurophysiology, 42:1297–315, 10 1979.

[105] I. Darian-Smith, K. O. Johnson, and R. W. Dykes.
"cold" fiber population innervating palmar and digital skin of the monkey: responses to cooling pulses.
Journal of neurophysiology, 36 2:325–46, 1973.

[106] R. H. LaMotte and J. N. Campbell.
Comparison of responses of warm and nociceptive c-fiber afferents in monkey with human judgments of thermal pain.

Journal of Neurophysiology, 41(2):509–528, 1978.

[107] K. O. Johnson, I. Darian-Smith, C. LaMotte, B. Johnson, and S. Oldfield.
Coding of incremental changes in skin temperature by a population of warm fibers in the monkey: correlation with intensity discrimination in man.
Journal of Neurophysiology, 42(5):1332–1353, 1979.

[108] F. Wang, E. Bélanger, S. L Côté, P. Desrosiers, S. A. Prescott, D. C. Côté, and Y. De Koninck.
Sensory afferents use different coding strategies for heat and cold.
Cell Rep, 23:2001–2013, 2018 May 15 2018.

[109] I. Darian-Smith and P. Kenins.
Innervation density of mechanoreceptive fibres supplying glabrous skin of the monkey's index finger.
The Journal of Physiology, 309(1):147–155, 1980.

[110] C E Connor, S S Hsiao, J R Phillips, and K O Johnson.
Tactile roughness: neural codes that account for psychophysical magnitude estimates.
The Journal of neuroscience : the official journal of the Society for Neuroscience, 10(12):3823–36, 1990.

[111] K. O. Johnson, S. Hsiao, and T. Yoshioka.
Neural coding and the basic law of psychophysics.
The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry, 8:111–21, 05 2002.

[112] J. R. Phillips and K. O. Johnson.
Tactile spatial resolution. II. Neural representation of Bars, edges, and gratings in monkey primary afferents.
Journal of Neurophysiology, 46(6):1192–1203, 2017.

[113] R. W. Van Boven and K. O. Johnson.
The limit of tactile spatial resolution in humans.
Neurology, 44(12):2361–2361, 1994.

[114] J. Tong, O. Mao, and D. Goldreich.
Two-point orientation discrimination versus the traditional two-point test for tactile spatial acuity assessment.
Frontiers in Human Neuroscience, 7:579, 2013.

[115] R. W. Van Boven, R. H. Hamilton, T. Kauffman, J. P. Keenan, and A. Pascual-Leone.
Tactile spatial resolution in blind braille readers.
Neurology, 54(12):2230–2236, 2000.

[116] A. Pruszynski and R. Johansson.
Edge-orientation processing in first-order tactile neurons.
Nature neuroscience, 17:1404–1410, 08 2014.

[117] K. Sathian, A. W. Goodwin, K. T. John, and I. Darian-Smith.
Perceived roughness of a grating: correlation with responses of mechanoreceptive afferents innervating the monkey's fingerpad.
The Journal of neuroscience : the official journal of the Society for Neuroscience, 9(4):1273–9, 1989.

[118] C. J. Cascio and K. Sathian.
Temporal cues contribute to tactile perception of roughness.
The Journal of neuroscience : the official journal of the Society for Neuroscience, 21(14):5289–96, 2001.

[119] A. W. Goodwin, K. T. John, K. Sathian, and I. Darian-Smith.
Spatial and temporal factors determining afferent fiber responses to a grating moving sinusoidally over the monkey's fingerpad.
The Journal of neuroscience : the official journal of the Society for Neuroscience, 9(4):1280–1293, 1989.

[120] E. M. Meftah, L. Belingard, and C. E. Chapman.
Relative effects of the spatial and temporal characteristics of scanned surfaces on human perception of tactile roughness using passive touch.
Experimental Brain Research, 132(3):351–361, 2000.

[121] J. J. Dicarlo, K. O. Johnson, and S. S. Hsiao.
Structure of receptive fields in area 3b of primary somatosensory cortex in the alert monkey.
Journal of Neuroscience, 18(7):2626–2645, 1998.

[122] E. P. Gardner, C. I. Palmer, H. A. Hamalainen, and S. Warren.
Simulation of motion on the skin. V. Effect of stimulus temporal frequency on the representation of moving bar patterns in primary somatosensory cortex of monkeys.
Journal of Neurophysiology, 67(1):37–63, 1992.

[123] F. Tremblay, S. A. Ageranioti-Belanger, and C. E. Chapman.
Cortical mechanisms underlying tactile discrimination in the monkey. i. role of primary somatosensory cortex in passive texture discrimination.
Journal of Neurophysiology, 76(5):3382–3403, 1996.

[124] W. Jiang, F. Tremblay, and C. E. Chapman.
Neuronal encoding of texture changes in the primary and the secondary somatosensory cortical areas of monkeys during passive texture discrimination.

Journal of Neurophysiology, 77(3):1656–1662, 1997.

[125] L. E. Krueger.
David katz's der aufbau der tastwelt (the world of touch): A synopsis.
Perception & Psychophysics, 7(6):337–341, Nov 1970.

[126] S. Bensmaïa and M. Hollins.
The vibrations of texture.
Somatosensory and Motor Research, 20(1):33–43, 2003.

[127] S. Bensmaïa, M. Hollins, and J. Yau.
Vibrotactile intensity and frequency information in the Pacinian system: A psychophysical model.
Perception and Psychophysics, 67(5):828–841, 2005.

[128] S. Bensmaïa and M. Hollins.
Pacinian representations of fine surface texture.
Perception and Psychophysics, 67(5):842–854, 2005.

[129] A. Weber, H. Saal, J. Lieber, JW. Cheng, L. Manfredi, J. Dammann, and S. Bensmaïa.
Spatial and temporal codes mediate the tactile perception of natural textures.
Proceedings of the National Academy of Sciences, 110(42):17107–17112, 2013.

[130] T. Yoshioka, J. C. Craig, G. C. Beck, and S. S. Hsiao.
Perceptual constancy of texture roughness in the tactile system.
Journal of Neuroscience, 31(48):17603–17611, 2011.

[131] Z. M. Boundy-Singer, H. Saal, and J. Bensmaïa.
Speed invariance of tactile texture perception.
Journal of Neurophysiology, 118(4):2371–2377, 2017.

[132] L. Manfredi, H. Saal, K. Brown, M. Zielinski, J. Dammann, V. Polashock, and S. Bensmaïa.
Natural scenes in tactile texture.
Journal of Neurophysiology, 111(9):1792–1802, 2014.

[133] H. Saal, X. Wang, and S. Bensmaïa.
Importance of spike timing in touch: an analogy with hearing?
Current Opinion in Neurobiology, 40:142 – 149, 2016.
Systems neuroscience.

[134] I. Jeffrey.
Photography : a concise history / Ian Jeffrey.
World of art. Thames and Hudson, London, 1981.

[135] L. D. Harmon.
Automated tactile sensing.
The International Journal of Robotics Research, 1(2):3–32, 1982.

[136] M. R. Cutkosky and W. Provancher.
Force and Tactile Sensing, pages 717–736.
Springer International Publishing, 2016.

[137] H. B. Muhammad, C. M. Oddo, L. Beccai, C. Recchiuto, C. J. Anthony, M. J. Adams, M. C. Carrozza, D. W. L. Hukins, and M. C. L. Ward.
Development of a bioinspired mems based capacitive tactile sensor for a robotic finger.
Sensors and Actuators A: Physical, 165(2):221 – 229, 2011.

[138] P. A. Schmidt, E. Ma, and R. P. W.
A sensor for dynamic tactile information with applications in human6robot interaction and object exploration.
Robotics and Autonomous Systems, 54(12):1005 – 1014, 2006.

[139] N. Jamali, M. Maggiali, F. Giovannini, G. Metta, and L. Natale.
A new design of a fingertip for the icub hand.
In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2705–2710, 2015.

[140] A. Schmitz, P. Maiolino, M. Maggiali, L. Natale, G. Cannata, and G. Metta.
Methods and technologies for the implementation of large-scale robot tactile sensors.
IEEE Transactions on Robotics, 27(3):389–400, June 2011.

[141] S. Chitta, J. Sturm, M. Piccoli, and W. Burgard.
Tactile sensing for mobile manipulation.
IEEE Transactions on Robotics, 27(3):558–568, June 2011.

[142] T. M. Huh, H. Choi, S. Willcox, S. Moon, and M. R. Cutkosky.
Dynamically reconfigurable tactile sensor for robotic manipulation.
IEEE Robotics and Automation Letters, 5(2):2562–2569, 2020.

[143] S. Stassi, V. Cauda, G. Canavese, and Candido F. Pirri.
Flexible tactile sensing based on piezoresistive composites: A review.
Sensors, 14(3):5296–5332, 2014.

[144] T. Shimizu, M. Shikida, K. Sato, and K. Itoigawa.
A new type of tactile sensor detecting contact force and hardness of an object.
In *Technical Digest. MEMS 2002 IEEE International Conference. Fifteenth IEEE International Conference on Micro Electro Mechanical Systems (Cat. No.02CH37266)*, pages 344–347, Jan 2002.

[145] A. Drimus, G. Kootstra, A. Bilberg, and D. Kragic.
Classification of rigid and deformable objects using a novel tactile sensor.
In *2011 15th International Conference on Advanced Robotics (ICAR)*, pages 427–434, June 2011.

[146] P. Silva Gir ao, P. Miguel P. Ramos, O. Postolache, and J. M. D. Pereira.
Tactile sensors for robotic applications.
Measurement, 46(3):1257 – 1271, 2013.

[147] J. Sinapov, V. Sukhoy, R. Sahai, and A. Stoytchev.
Vibrotactile recognition and categorization of surfaces by a humanoid robot.
IEEE Transactions on Robotics, 27(3):488–497, June 2011.

[148] M. Ahmed, M. M. Chitteboyina, D. P. Butler, and Z. 7elik-Butler.
Mems force sensor in a flexible substrate using nichrome piezoresistors.
IEEE Sensors Journal, 13(10):4081–4089, Oct 2013.

[149] C. M. Oddo, M. Controzzi, L. Beccai, C. Cipriani, and M. C. Carrozza.
Roughness encoding for discrimination of surfaces in artificial active-touch.
IEEE Transactions on Robotics, 27(3):522–533, June 2011.

[150] K. Noda, H. Onoe, E. Iwase, K. Matsumoto, and I. Shimoyama.
Flexible tactile sensor for shear stress measurement using transferred sub-Îijm-thick si piezoresistive cantilevers.
Journal of Micromechanics and Microengineering, 22(11), November 2012.

[151] F. Vidal-Veru, O. Oballe-Peinado, Jose A. SÃąnchez-Duran, J. Castellanos-Ramos, and R. Navas-Gonzalez.
Three realizations and comparison of hardware for piezoresistive tactile sensors.
Sensors (Basel, Switzerland), 11:3249–66, 12 2011.

[152] B. Jaffe.
Piezoelectric ceramics.
Journal of the American Ceramic Society, 41(11):494–498, 1958.

[153] Z. Yi, Y. Zhang, and J. Peters.
Bioinspired tactile sensor for surface roughness discrimination.
Sensors and Actuators A: Physical, 255:46 – 53, 2017.

[154] Z. Yi and Y. Zhang.
Bio-inspired tactile fa-i spiking generation under sinusoidal stimuli.
Journal of Bionic Engineering, 13(4):612 – 621, 2016.

[155] D. Goger, N. Gorges, and H. Worn.
Tactile sensing for an anthropomorphic robotic hand: Hardware and signal processing.
In *2009 IEEE International Conference on Robotics and Automation*, pages 895–901, May 2009.

[156] L. Seminara, M. Capurro, P. Cirillo, G. Cannata, and M. Valle.
Electromechanical characterization of piezoelectric pvdf polymer films for tactile sensors in robotics applications.
Sensors and Actuators A: Physical, 169(1):49 – 58, 2011.

[157] C. Li, P. Wu, S. Lee, A. Gorton, M. J. Schulz, and C. H. Ahn.
Flexible dome and bump shape piezoelectric tactile sensors using pvdf-trfe copolymer.
Journal of Microelectromechanical Systems, 17(2):334–341, April 2008.

[158] S. Sokhanvar, M. Packirisamy, and J. Dargahi.
A multifunctional PVDF-based tactile sensor for minimally invasive surgery.
Smart Materials and Structures, 16(4):989–998, jun 2007.

[159] K. Takashima, S. Horie, T. Mukai, K. Ishida, and K. Matsushige.
Piezoelectric properties of vinylidene fluoride oligomer for use in medical tactile sensor applications.
Sensors and Actuators A: Physical, 144(1):90 – 96, 2008.

[160] A. Bonakdar and N. Narayanan.
Determination of tissue properties using microfabricated piezoelectric tactile sensor during minimally invasive surgery.
Sensor Review, 30:233–241, 06 2010.

[161] M. W. Strohmayr and D. Schneider.
The dlr artificial skin step ii: Scalability as a prerequisite for whole-body covers.
In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4721–4728, Nov 2013.

[162] M. L. Hammock, A. Chortos, B. C. K. Tee, J. B. H. Tok, and Z. Bao.
25th anniversary article: The evolution of electronic skin (e-skin): A brief history, design considerations, and recent progress.
Advanced Materials, 25(42):5997–6038, 2013.

[163] R. S. Dahiya, D. Cattin, A. Adami, C. Collini, L. Barboni, M. Valle, L. Lorenzelli, R. Oboe, G. Metta, and F. Brunetti.
Towards tactile sensing system on chip for robotic applications.
IEEE Sensors Journal, 11(12):3216–3226, Dec 2011.

[164] P. Piacenza, S. Sherman, and M. Ciocarlie.
Data-driven super-resolution on a tactile dome.
IEEE Robotics and Automation Letters, 3(3):1434–1441, July 2018.

[165] Y. Tenzer, L. P. Jentoft, and R. D. Howe.
The feel of mems barometers: Inexpensive and easily customized tactile array sensors.
IEEE Robotics Automation Magazine, 21(3):89–95, Sep. 2014.

[166] L. U. Odhner, L. P. Jentoft, M. R. Claffee, N. Corson, Y. Tenzer, R. R. Ma, M. Buehler, R. Kohout, R. D. Howe, and A. M. Dollar.
A compliant, underactuated hand for robust manipulation.
The International Journal of Robotics Research, 33(5):736–752, 2014.

[167] L. P. Jentoft, Y. Tenzer, D. Vogt, Jia Liu, R. J. Wood, and R. D. Howe.
Flexible, stretchable tactile arrays from mems barometers.
In *2013 16th International Conference on Advanced Robotics (ICAR)*, pages 1–6, Nov 2013.

[168] P. Kampmann and F. Kirchner.
Integration of fiber-optic sensor arrays into a multi-modal tactile sensor processing system for robotic end-effectors.
Sensors, 14(4):6854–6876, 2014.

[169] H. Xie, A. Jiang, L. Seneviratne, and K. Althoefer.
Pixel-based optical fiber tactile force sensor for robot manipulation.
Sensor, pages 1–4, Oct 2012.

[170] H. Maekawa, K. Tanie, K. Komoriya, M. Kaneko, C. Horiguchi, and T. Sugawara.
Development of a finger-shaped tactile sensor and its evaluation by active touch.
In *Proceedings 1992 IEEE International Conference on Robotics and Automation*, pages 1327–1334 vol.2, May 1992.

[171] G. Obinata, A. Dutta, N. Watanabe, and N. Moriyama.
Vision Based Tactile Sensor Using Transparent Elastic Fingertip for Dexterous Handling.
02 2007.

[172] T. Corradi, P. Hall, and P. Iravani.
5 2015.
IEEE Interational Conference on Robotics and Automation (ICRA) 2015 ; Conference date: 26-05-2015 Through 30-05-2015.

[173] I. Huang, J. Liu, and R. Bajcsy.
A depth camera-based soft fingertip device for contact region estimation and perception-action coupling.

In *2019 International Conference on Robotics and Automation (ICRA)*, pages 8443–8449, May 2019.

[174] M. K. Johnson and E. H. Adelson.
Retrographic sensing for the measurement of surface texture and shape.
In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1070–1077, June 2009.

[175] M. K. Johnson, F. Cole, A. Raj, and E. H. Adelson.
Microgeometry capture using an elastomeric sensor.
ACM Transactions on Graphics (Proc. ACM SIGGRAPH), 30(4):46:1–46:8, 2011.

[176] Robert J. Woodham.
Photometric Method For Determining Surface Orientation From Multiple Images.
Optical Engineering, 19(1):139 – 144, 1980.

[177] E. Donlon, S. Dong, M. Liu, J. Li, E. Adelson, and A. Rodríguez.
Gelslim: A high-resolution, compact, robust, and calibrated tactile-sensing finger.
2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 1927–1934, 2018.

[178] A. Church, J. W. James, L. Cramphorn, and N. F. Lepora.
Tactile model O: fabrication and testing of a 3d-printed, three-fingered tactile robot hand.
CoRR, abs/1907.07535, 2019.

[179] B. Winstone, G. Griffiths, T. Pipe, C. Melhuish, and J. Rossiter.
Tactip - tactile fingertip device, texture analysis through optical tracking of skin features.
In Nathan F. Lepora, Anna Mura, Holger G. Krapp, Paul F. M. J. Verschure, and Tony J. Prescott, editors, *Biomimetic and Biohybrid Systems*, pages 323–334, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[180] H. Khamis, B. Xia, and S. J. Redmond.
A novel optical 3d force and displacement sensor âĂŞ towards instrumenting the papillarray tactile sensor.
Sensors and Actuators A: Physical, 291:174 – 187, 2019.

[181] N. Wettels, V. J. Santos, R. S. Johansson, and G. E. Loeb.
Biomimetic tactile sensor array.
Advanced Robotics, 22(8):829–849, 2008.

[182] SynTouch, LLC.
BioTac product manual, 2019.

[183] J. Fishel and G. Loeb.
Bayesian exploration for intelligent identification of textures.
Frontiers in neurorobotics, 6:4, 06 2012.

[184] P. Mittendorfer and G. Cheng.
Humanoid multimodal tactile-sensing modules.
IEEE Transactions on robotics, 27:401–10, 2011.

[185] J. Engel, J. Chen, Z. Fan, and C. Liu.
Polymer micromachined multimodal tactile sensors.
Sensors and Actuators A: Physical, 117(1):50 – 61, 2005.

[186] A. Cretu A. de Oliveira and E. Petriu.
Multimodal bio-inspired tactile sensing module for surface characterization.
Sensors, 17(6), 2017.

[187] E. Knoop and J. Rossiter.
Dual-mode compliant optical tactile sensor.
In *2013 IEEE International Conference on Robotics and Automation*, pages 1006–1011, 2013.

[188] C. Chorley, C. Melhuish, T. Pipe, and J. Rossiter.
Tactile edge detection.
In *2010 IEEE Sensors*, pages 2593–2598, 2010.

[189] N. F. Lepora, K. Aquilina, and L. Cramphorn.
Exploratory tactile servoing with active touch.
IEEE Robotics and Automation Letters, 2(2):1156–1163, 2017.

[190] B. Ward-Cherrier, N. Rojas, and N. F. Lepora.
Model-free precise in-hand manipulation with a 3d-printed tactile gripper.
IEEE Robotics and Automation Letters, 2(4):2056–2063, 2017.

[191] B. Winstone, C. Melhuish, T. Pipe, M. Callaway, and S. Dogramadzi.
Toward bio-inspired tactile sensing capsule endoscopy for detection of submucosal tumors.
IEEE Sensors Journal, 17:848–857, 2016.

[192] R. Calandra, A. Owens, D. Jayaraman, J. Lin, Y. Wenzhen, J. Malik, E. H. Adelson, and S. Levine.
More than a feeling: Learning to grasp and regrasp using vision and touch.
CoRR, abs/1805.11085, 2018.

[193] R. Li and E. H. Adelson.

Sensing and recognizing surface textures using a gelsight sensor.
In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1241–1247, June 2013.

[194] Y. Chebotar, K. Hausman, Z. Su, G. S. Sukhatme, and S. Schaal.
Self-supervised regrasping using spatio-temporal tactile features and reinforcement learning.
In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1960–1966, Oct 2016.

[195] J. Sinapov, T. Bergquist, C. Schenck, S. Ohiri, U.and Griffith, and A. Stoytchev.
Interactive object recognition using proprioceptive and auditory feedback.
Int. J. Rob. Res., 30(10):1250–1262, September 2011.

[196] K. Aquilina, D. A. W. Barton, and N. F. Lepora.
Shear-invariant sliding contact perception with a soft tactile sensor.
In *2019 International Conference on Robotics and Automation (ICRA)*, pages 4283–4289, May 2019.

[197] D. Driess, D. Hennes, and M. Toussaint.
Active multi-contact continuous tactile exploration with gaussian process differential entropy.
In *2019 International Conference on Robotics and Automation (ICRA)*, pages 7844–7850, May 2019.

[198] E. Roscow, C. Kent, U. Leonards, and N. F. Lepora.
Discrimination-based perception for robot touch.
In Nathan F. Lepora, Anna Mura, Michael Mangan, Paul F.M.J. Verschure, Marc Desmulliez, and Tony J. Prescott, editors, *Biomimetic and Biohybrid Systems*, pages 498–502, Cham, 2016. Springer International Publishing.

[199] L. Cramphorn, B. Ward-Cherrier, and N. F. Lepora.
Addition of a biomimetic fingerprint on an artificial fingertip enhances tactile spatial acuity.
IEEE Robotics and Automation Letters, 2(3):1336–1343, 2017.

[200] N. Gorges, S. Escaida Navarro, D. G6ger, and H. W6rn.
Haptic object recognition using passive joints and haptic key features.
In *2010 IEEE International Conference on Robotics and Automation*, pages 2349–2355, May 2010.

[201] H. Liu, J. Greco, X. Song, J. Bimbo, L. Seneviratne, and K. Althoefer.

Tactile image based contact shape recognition using neural network.
In *2012 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, pages 138–143, Sep. 2012.

[202] S. K. Yeung, E. M. Petriu, W. S. McMath, and D. C. Petriu.
High sampling resolution tactile sensor for object recognition.
IEEE Transactions on Instrumentation and Measurement, 43(2):277–282, April 1994.

[203] H. Liu, S. Luo, and K. Althoefer.
Novel tactile-sift descriptor for object shape recognition.
IEEE Sensors Journal, 15, 09 2015.

[204] A. Schneider, J. Sturm, C. Stachniss, M. Reisert, H. Burkhardt, and W. Burgard.
Object identification with tactile sensors using bag-of-features.
In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 243–248, Oct 2009.

[205] Yi-Hung Liu, Yu-Tsung Hsiao, Wei-Teng Cheng, Yan-Chen Liu, and Jui-Yiao Su.
Low-resolution tactile image recognition for automated robotic assembly using kernel pca-based feature fusion and multiple kernel learning-based support vector machine.
Mathematical Problems in Engineering, 2014:1–11, 02 2014.

[206] J. Kwiatkowski, D. Cockburn, and V. Duchaine.
Grasp stability assessment through the fusion of proprioception and tactile signals using convolutional neural networks.
In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 286–292, Sept 2017.

[207] D. Cockburn, J. Roberge, T. Le, A. Maslyczyk, and V. Duchaine.
Grasp stability assessment through unsupervised feature learning of tactile images.
In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2238–2244, May 2017.

[208] M. Bauza, O. Canal, and A. Rodriguez.
Tactile mapping and localization from high-resolution tactile imprints.
pages 3811–3817, 05 2019.

[209] S. Gupta, L. Lorenzelli, and R. Dahiya.
Multifunctional flexible pvdf-trfe/batio3 based tactile sensor for touch and temperature monitoring.
In *2017 IEEE Sensors*, pages 1–3, Oct 2017.

[210] Z. Su, J. Fishel, T. Yamamoto, and G. Loeb.
Use of tactile feedback to control exploratory movements to characterize object compliance.
Frontiers in neurorobotics, 6:7, 07 2012.

[211] Z. Kappassov, D. Baimukashev, O. Adiyatov, S. Salakchinov, Y. Massalin, and H. A. Varol.
A series elastic tactile sensing array for tactile exploration of deformable and rigid objects.
In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 520–525, Oct 2018.

[212] W. Yuan, M. A. Srinivasan, and E. H. Adelson.
Estimating object hardness with a gelsight touch sensor.
In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 208–215, Oct 2016.

[213] Z. Kappassov, J.-A. Corrales, and V. Perdereau.
Touch driven controller and tactile features for physical interactions.
Robotics and Autonomous Systems, 123:103332, 2020.

[214] R. Li, R. Platt, W. Yuan, A. ten Pas, N. Roscup, M. A. Srinivasan, and E. Adelson.
Localization and manipulation of small parts using gelsight tactile sensing.
In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3988–3993, Sep. 2014.

[215] N. Pestell, L. Cramphorn, F. Papadopoulos, and N. F. Lepora.
A sense of touch for the shadow modular grasper.
IEEE Robotics and Automation Letters, 4(2):2220–2226, April 2019.

[216] Y. Mukaibo, H. Shirado, M. Konyo, and T. Maeno.
Development of a texture sensor emulating the tissue structure and perceptual mechanism of human fingers.
In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pages 2565–2570, April 2005.

[217] J. Edwards, J. Lawry, J. Rossiter, and C. Melhuish.
Extracting textural features from tactile sensors.
Bioinspiration & Biomimetics, 3(3):035002, jun 2008.

[218] M. Johnsson and C. Balkenius.
Sense of touch in robots with self-organizing maps.
IEEE Transactions on Robotics, 27(3):498–507, June 2011.

[219] V. A. Ho, T. Araki, M. Makikawa, and S. Hirai.

Experimental investigation of surface identification ability of a low-profile fabric tactile sensor.
In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4497–4504, Oct 2012.

[220] S. Bensmaïa and M. Hollins.
Pacinian representations of fine surface texture.
Perception & Psychophysics, 67(5):842–854, Jul 2005.

[221] M. Hollins, R. Faldowski, S. Rao, and F. Young.
Perceptual dimensions of tactile surface texture: A multidimensional scaling analysis.
Perception & Psychophysics, 54(6):697–705, Nov 1993.

[222] P. Dallaire, P. Gigu8re, D. 9mond, and B. Chaib-draa.
Autonomous tactile perception: A combined improved sensing and bayesian nonparametric approach.
Robotics and Autonomous Systems, 62(4):422 – 435, 2014.

[223] R. Li and E. Adelson.
Sensing and recognizing surface textures using a gelsight sensor.
pages 1241–1247, 06 2013.

[224] T. Ojala, M. Pietikainen, and T. Maenpaa.
Multiresolution gray-scale and rotation invariant texture classification with local binary patterns.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7):971–987, July 2002.

[225] S. Luo, W. Yuan, E. Adelson, A. Cohn, and R. Fuentes.
Vitac: Feature sharing between vision and tactile sensing for cloth texture recognition.
01 2018.

[226] A. Krizhevsky, I. Sutskever, and G. Hinton.
Imagenet classification with deep convolutional neural networks.
Neural Information Processing Systems, 25, 01 2012.

[227] Wenzhen Yuan, Yuchen Mo, Shaoxiong Wang, and Edward Adelson.
Active clothing material perception using tactile sensing and deep learning.
In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–8, 05 2018.

[228] S. S. Baishya and B. B4uml.
Robust material classification with a tactile skin using deep learning.

In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8–15, Oct 2016.

[229] T. Taunyazov, H. F. Koh, Y. Wu, C. Cai, and H. Soh.
Towards effective tactile identification of textures using a hybrid touch approach.
In *2019 International Conference on Robotics and Automation (ICRA)*, pages 4269–4275, May 2019.

[230] J. M. Romano and K. J. Kuchenbecker.
Methods for robotic tool-mediated haptic surface recognition.
In *2014 IEEE Haptics Symposium (HAPTICS)*, pages 49–56, Feb 2014.

[231] H. Kawasaki, T. Komatsu, K. Uchiyama, and T. Kurimoto.
Dexterous anthropomorphic robot hand with distributed tactile sensor: Gifu hand ii.
In *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.99CH37028)*, volume 2, pages 782–787 vol.2, Oct 1999.

[232] Y. Su, Y. Wu, K. Lee, Z. Du, and Y. Demiris.
Robust grasping for an under-actuated anthropomorphic hand under object position uncertainty.
In *2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012)*, pages 719–725, Nov 2012.

[233] J. Bimbo, S. Luo, K. Althoefer, and H. Liu.
In-hand object pose estimation using covariance-based tactile to geometry matching.
IEEE Robotics and Automation Letters, 1(1):570–577, Jan 2016.

[234] J. Bimbo, P. Kormushev, K. Althoefer, and H. Liu.
Global estimation of an object2s pose using tactile sensing.
Advanced Robotics, 29(5):363–374, 2015.

[235] A. Vasquez, Z. Kappasov, and V. Perdereau.
In-hand object shape identification using invariant proprioceptive signatures.
2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 965–970, 2016.

[236] A. J. Spiers, M. V. Liarokapis, B. Calli, and A. M. Dollar.
Single-grasp object classification and feature extraction with simple robot hands and tactile sensors.
IEEE Transactions on Haptics, 9(2):207–220, April 2016.

[237] J. Antonio R. Corrales, V. Perdereau, and F. Medina.

Multi-fingered robotic hand planner for object reconfiguration through a rolling contact evolution model.
pages 625–630, 05 2013.

[238] H. Yussof, M. Ohka, H. Suzuki, and N. Morisawa.
Tactile Sensing-based Control System for Dexterous Robot Manipulation, pages 199–213.
Springer Netherlands, Dordrecht, 2009.

[239] W. Chen, H. Khamis, I. Birznieks, N. F. Lepora, and S. J. Redmond.
Tactile sensors for friction estimation and incipient slip detection toward dexterous robotic manipulation: A review.
IEEE Sensors Journal, 18(22):9049–9064, Nov 2018.

[240] V. A. Ho and S. Hirai.
Understanding slip perception of soft fingertips by modeling and simulating stick-slip phenomenon.
In *Robotics: Science and Systems*, 2011.

[241] J. Jockusch, J. Walter, and H. Ritter.
A tactile sensor system for a three-fingered robot manipulator.
In *Proceedings of International Conference on Robotics and Automation*, volume 4, pages 3080–3086 vol.4, April 1997.

[242] S. Shirafuji and K. Hosoda.
Detection and prevention of slip using sensors with different properties embedded in elastic artificial skin on the basis of previous experience.
Robotics and Autonomous Systems, 62(1):46 – 52, 2014.
New Boundaries of Robotics.

[243] S. Dong, D. Ma, E. Donlon, and A. Rodriguez.
Maintaining grasps within slipping bounds by monitoring incipient slip.
In *2019 International Conference on Robotics and Automation (ICRA)*, pages 3818–3824, May 2019.

[244] Z. Su, K. Hausman, Y. Chebotar, A. Molchanov, G. E. Loeb, G. S. Sukhatme, and S. Schaal.
Force estimation and slip detection/classification for grip control using a biomimetic tactile sensor.
In *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*, pages 297–303, Nov 2015.

[245] M. Andrychowicz et al.
Learning dexterous in-hand manipulation.
The International Journal of Robotics Research, page 027836491988744, 11 2019.

[246] Avago Technologies.
ADNS-3080, 2009.

[247] L. Cramphorn, J. Lloyd, and N. F. Lepora.
Voronoi features for tactile sensing: Direct inference of pressure, shear, and contact locations.
In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018.

[248] N. F. Lepora.
Biomimetic active touch with fingertips and whiskers.
IEEE Transactions on Haptics, 9(2):170–183, 2016.

[249] B. Ward-Cherrier, N. Pestell, L. Cramphorn, B. Winstone, M. E. Giannaccini, J. Rossiter, and N. F. Lepora.
The tactip family: Soft optical tactile sensors with 3d-printed biomimetic morphologies.
Soft Robotics, 2018.

[250] D. Font, M. Tresanchez, T. Pallejá, M. Teixidó, and J. Palacín.
Characterization of a low-cost optical flow sensor when using an external laser as a direct illumination source.
Sensors, 11:11856–70, 2011.

[251] B. G. Lipták.
Instrument Engineers' Handbook, Volume Two: Process Control and Optimization.
CRC/Taylor & Francis, 2006.

[252] Chris Liechti pyserial github page.
`https://github.com/pyserial/pyserial`.
Accessed: 2019-06-25.

[253] Nick Pestell adns3080 page.
`https://github.com/nick-pestell/adns3080`.
Accessed: 2020-04-10.

[254] Michael Dawson-Haggerty open_abb github page.
`https://github.com/robotics/open_abb`.
Accessed: 2019-04-17.

[255] Nick Pestell dual-modal tactile perception and exploration video.
`https://www.youtube.com/watch?v=_4UjTWtyK5w`.
Accessed: 2020-04-13.

[256] N. Wettels, J. A. Fishel, and G. E. Loeb.

Multimodal tactile sensor.
In *The Human Hand as an Inspiration for Robot Hand Development*, pages 405–429. Springer, 2014.

[257] J. Segil, R. Patel, J. Klingner, R. F. Weir, and N. Correll.
Multi-modal prosthetic fingertip sensor with proximity, contact, and force localization capabilities.
Advances in Mechanical Engineering, 11(4):1687814019844643, 2019.

[258] M. S. Arian, C. A. Blaine, G. E. Loeb, and J. A. Fishel.
Using the biotac as a tumor localization tool.
In *2014 IEEE Haptics Symposium (HAPTICS)*, pages 443–448, 2014.

[259] J. A. Fishel and G. E. Loeb.
Sensing tactile microvibrations with the biotac âĂŤ comparison with human sensitivity.
In *2012 4th IEEE RAS EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, pages 1122–1127, June 2012.

[260] Z. Abderrahmane, G. Ganesh, A. Crosnier, and A. Cherubini.
Haptic zero-shot learning: Recognition of objects never touched before.
Robotics and Autonomous Systems, 105:11 – 25, 2018.

[261] E. Dean-Leon, F. Bergner, K. Ramirez-Amaro, and G. Cheng.
From multi-modal tactile signals to a compliant control.
In *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, pages 892–898, 2016.

[262] J. M. Romano, K. Hsiao, G. Niemeyer, S. Chitta, and K. J. Kuchenbecker.
Human-inspired robotic grasp control with tactile sensing.
IEEE Transactions on Robotics, 27(6):1067–1079, Dec 2011.

[263] E. Kim, S. Wellnitz, S. Ciupek, E. Lumpkin, and . Gerling.
Force sensor in simulated skin and neural model mimic tactile sai afferent spiking response to ramp and hold stimuli.
Journal of neuroengineering and rehabilitation, 9:45, 07 2012.

[264] N. Salimi-Nezhad, E. Ilbeigi, M. Amiri, E. Falotico, and C. Laschi.
A digital hardware system for spiking network of tactile afferents.
Frontiers in Neuroscience, 13:1330, 2020.

[265] W. W. Lee, S. L. Kukreja, and N. V. Thakor.
Discrimination of dynamic tactile contact by temporally precise event sensing in spiking neuromorphic networks.

Frontiers in Neuroscience, 11:5, 2017.

[266] I. Birznieks, V. G. Macefield, G. Westling, and R. S. Johansson.
Slowly adapting mechanoreceptors in the borders of the human fingernail encode fingertip forces.
Journal of Neuroscience, 29(29):9370–9379, 2009.

[267] Kingstate, Taiwan.
KECG3642TF-A, 2019.

[268] Fredrik Portstrom python-v4l2capture github page.
`https://github.com/gebart/python-v4l2capture`.
Accessed: 2019-10-09.

[269] Matthias Geier python-sounddevice github page.
`https://github.com/spatialaudio/python-sounddevice`.
Accessed: 2019-10-09.

[270] J. James, N. Pestell, and N. Lepora.
Slip detection with a biomimetic tactile sensor.
IEEE Robotics and Automation Letters, 3(4):1–1, 07 2018.

[271] I. Birznieks, P. Jenmalm, A. Goodwin, and R. Johansson.
Encoding of direction of fingertip forces by human tactile afferents.
The Journal of neuroscience : the official journal of the Society for Neuroscience, 21:8222–37, 11 2001.

[272] S. Bensmaïa.
A transduction model of the meissner corpuscle.
Mathematical Biosciences, 176(2):203 – 217, 2002.

[273] S. S. Kim, S. Sripati, and S. Bensmaïa.
Predicting the timing of spikes evoked by tactile stimulation of the hand.
Journal of neurophysiology, 104:1484–96, 09 2010.

[274] G. Bradski.
The OpenCV Library.
Dr. Dobb's Journal of Software Tools, 2000.

[275] A. W. Goodwin and J. W. Morley.
Sinusoidal movement of a grating across the monkey's fingerpad: representation of grating and movement features in afferent fiber responses.
Journal of Neuroscience, 7(7):2168–2180, 1987.

[276] A. B. Vallbo, K. A. Olsson, K.-G. Westberg, and F. J. Clark.
Microstimulation of Single Tactile Afferents from the Human Hand: Sensory Attributes Related to Unit Type and Properties of Receptive Fields.
Brain, 107(3):727–749, 09 1984.

[277] J. Platkiewicz, H. Lipson, and V. Hayward.
Haptic Edge Detection Through Shear.
Scientific Reports, 6:23551, March 2016.

[278] Kiran Dandekar, Balasundar Raju, and Mandayam Srinivasan.
3-d finite-element models of human and monkey fingertips to investigate the mechanics of tactile sense.
Journal of biomechanical engineering, 125:682–91, 11 2003.

[279] Z. Del Prete, S. P. Baker, and P. Grigg.
Stretch responses of cutaneous ra afferent neurons in mouse hairy skin.
Journal of Neurophysiology, 89(3):1649–1659, 2003.
PMID: 12612035.

[280] P. S. Khalsa, A. H. Hoffman, and P. Grigg.
Mechanical states encoded by stretch-sensitive neurons in feline joint capsule.
Journal of Neurophysiology, 76(1):175–187, 1996.
PMID: 8836217.

[281] P. S. Khalsa, R. H. Lamotte, and P. Grigg.
Tensile and compressive responses of nociceptors in rat hairy skin.
Journal of Neurophysiology, 78(1):492–505, 1997.
PMID: 9242296.

[282] J. R. Phillips and K. O. Johnson.
Tactile spatial resolution. iii. a continuum mechanics model of skin predicting mechanoreceptor responses to bars, edges, and gratings.
Journal of neurophysiology, 46(6):1204âĂŤ1225, December 1981.

[283] L. Cramphorn, B. Ward-Cherrier, and N. F. Lepora.
Tactile manipulation with biomimetic active touch.
In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 123–129, 2016.

[284] B. Ward-Cherrier, N. Pestell, and N. F. Lepora.
Neurotac: A neuromorphic optical tactile sensor applied to texture recognition, 2020.

[285] A. Lenail.
NN-SVG: Publication-ready neural network architecture schematics.
Journal of Open Source Software, 4(33):747, 2019.

[286] F. Chollet et al.
Keras.
`https://keras.io`, 2015.

[287] M. Abadi et al.
TensorFlow: Large-scale machine learning on heterogeneous systems, 2015.
Software available from tensorflow.org.

[288] R. Klatzky and C. L. Reed.
Haptic exploration.
Scholarpedia, 4(8):7941, 2009.

[289] Y. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller.
Efficient backprop.
In *Neural Networks: Tricks of the Trade, This Book is an Outgrowth of a 1996 NIPS Workshop*, page 9650, Berlin, Heidelberg, 1998. Springer-Verlag.

[290] L. Perez and J. Wang.
The effectiveness of data augmentation in image classification using deep learning.
CoRR, abs/1712.04621, 2017.

[291] S. Washburn M. Hollins, S. J. Bensma.
Vibrotactile adaptation impairs discrimination of fine, but not coarse, textures.
Somatosensory & Motor Research, 18(4):253–262, 2001.

[292] L. Skedung, M. Arvidsson, J. Young Chung, C. Stafford, B. Berglund, and M. Rutland.
Feeling small: Exploring the tactile perception limits.
3:2617, 2013.

[293] R. S. Johansson and G. Westling.
Roles of glabrous skin receptors and sensorimotor memory in automatic control of precision grip when lifting rougher or more slippery objects.
Experimental Brain Research, 56(3):550–564, Oct 1984.

[294] E. Hyttinen, D. Kragic, and R. Detry.
Learning the tactile signatures of prototypical object parts for robust part-based grasping of novel objects.
In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4927–4932, May 2015.

[295] L. Pinto and A. Gupta.
Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours.
In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3406–3413, May 2016.

[296] M. Li, K. Hang, D. Kragic, and A. Billard.
Dexterous grasping under shape uncertainty.
Robotics and Autonomous Systems, 75:352 – 364, 2016.

[297] E. Jones, T. Oliphant, and P. Peterson et al.
SciPy: Open source scientific tools for Python, 2001–.
[Online; accessed <today>].

[298] F. et al. Pedregosa.
Scikit-learn: Machine learning in Python.
Journal of Machine Learning Research, 12:2825–2830, 2011.

[299] B. Calli, A. Walsman, A. Singh, S. Srinivasa, P. Abbeel, and A. M. Dollar.
Benchmarking in manipulation research: Using the yale-cmu-berkeley object and model set.
IEEE Robotics Automation Magazine, 22(3):36–52, Sep. 2015.

[300] Nick Pestell a sense of touch for the shadow modular grasper video.
`https://www.youtube.com/watch?v=urHs56hqpcM`.
Accessed: 2020-04-10.

[301] M. A. Roa and R. Suarez.
Grasp quality measures: Review and performance.
Autonomous Robots, 38:65–88, 07 2014.

[302] W. Maass.
Networks of spiking neurons: The third generation of neural network models.
Neural Networks, 10(9):1659 – 1671, 1997.

[303] FTDI ftdi website.
`https://www.ftdichip.com/index.html`.
Accessed: 2019-04-17.